Neural Computing and Applications (2023) 35:20463–20568
https://doi.org/10.1007/s00521-023-08863-9

REVIEW

A perspective on human activity recognition from inertial motion data

Walid Gomaa[1,2] · Mohamed A. Khamis[1,3]

Received: 20 January 2023 / Accepted: 5 July 2023 / Published online: 31 July 2023

Abstract

Human activity recognition (HAR) using inertial motion data has gained a lot of momentum in recent years both in research and industrial applications. From the abstract perspective, this has been driven by the rapid dynamics for building intelligent, smart environments, and ubiquitous systems that cover all aspects of human life including healthcare, sports, manufacturing, commerce, etc., which necessitate and subsume activity recognition aiming at recognizing the actions, characteristics, and goals of one or more agent(s) from a temporal series of observations streamed from one or more sensors. From a more concrete and seemingly orthogonal perspective, such momentum has been driven by the ubiquity of inertial motion sensors on-board mobile and wearable devices including smartphones, smartwatches, etc. In this paper we give an introductory and a comprehensive survey to the subject from a given perspective. We focus on a subset of topics, that we think are major, that will have significant and influential impacts on the future research and industrial-scale deployment of HAR systems. These include: (1) a comprehensive and detailed description of the inertial motion benchmark datasets that are publicly available and/or accessible, (2) feature selection and extraction techniques and the corresponding learning methods used to build workable HAR systems; we survey classical handcrafted datasets as well as data-oriented automatic representation learning approach to the subject, (3) transfer learning as a way to overcome many hurdles in actual deployments of HAR systems on a large scale, (4) embedded implementations of HAR systems on mobile and/or wearable devices, and finally (5) we touch on adversarial attacks, a topic that is essentially related to the security and privacy of HAR systems. As the field is very huge and diverse, this article is by no means comprehensive; it is though meant to provide a logically and conceptually rather complete picture to advanced practitioners, as well as to present a readable guided introduction to newcomers. Our logical and conceptual perspectives mimic the typical data science pipeline for state-of-the-art AI-based systems.

Keywords Human activity recognition · Inertial measurement unit · Transfer learning · Feature selection and extraction · Adversarial attacks · Machine learning · Smartwatches · Wearable devices · Mobile devices · Embedded implementations · Biometrics · Healthcare

1 Introduction

Human-centered computing is an emerging research and application area that aims at understanding the human behavior (and psychology) and integrating users and the social context with the digital technology. This necessitates and subsumes *human activity recognition* (HAR) which aims at recognizing the actions, characteristics, and goals of one or more individual(s) from a temporal series of observations streamed from one or more sensors. Automatic recognition of behavioral context can serve in many domains including health monitoring, elderly care, indoor and outdoor sporting, virtual coaching, surveillance systems, robotics learning and cooperation, smart homes, Industry 4.0, predictive maintenance, human–robot

Walid Gomaa, Mohamed A. Khamis have contributed equally to this work.

✉ Mohamed A. Khamis
mohamed.khamis@ejust.edu.eg; maomar@ejada.com

Walid Gomaa
walid.gomaa@ejust.edu.eg

[1] Cyber-Physical Systems Lab, Egypt-Japan University of Science and Technology (E-JUST), New Borg El-Arab City, Alexandria 21934, Egypt

[2] Faculty of Engineering, Alexandria University, Alexandria 21544, Egypt

[3] AI Unit, Ejada Systems Ltd., Alexandria, Egypt

interaction, the military, etc. A comprehensive survey of applications can be found in [1].

Generally, the objectives of HAR systems can be summarized as follows: (1) determining (both in real-time and offline) the ongoing actions/activities of a person, a group of persons, or even the crowd based on sensory observation data, (2) assessment of the ongoing action, for example, if the action is an indoor squat exercise, what is the performance level of the player and what mistakes done during exercising, (3) determining certain characteristics of the subject(s) such as the identity of people in the given space, biometrics such gender, age, weight, etc., (4) knowledge about the context within which the observed activities have been taking place, and (5) estimating the body weight.

For our purposes in this article, we consider an action or activity as an ordered intelligible sequence of bodily movements.

HAR systems can be classified based on the modality of sensory information used, as the kind of sensory data greatly affects the types of features, algorithms, architectures, and techniques used for analysis. Generally, we can identify the following streams of research and developments in HAR systems: (1) HAR systems based on visual data (such as images and video streams), (2) HAR systems based on inertial motion streaming from sensors such as IMUs (Inertial Measurement Units), (3) HAR systems based on the received signal strength (RSS) from commodity routers installed in the surrounding environment [2, 3], (4) HAR systems based on infrared signals, and (5) HAR systems based on audio signals.

To the best of our knowledge, there is not much work fusing several of these modalities together. In the current study, we focus exclusively on the second modality, namely, inertial motion timeseries data. This type of data has the advantage of being more intrinsic in capturing the motion dynamics of the activity, and is not affected by any irregularities in the surrounding environment such as the absence of light for visual capturing. However, this is also its weakness as it has lower capability, in contrast to the other modalities, in capturing the contextual environment, and interactions with other objects and/or persons in that environment.

The idea of an *automatic HAR system* depends on collecting measurements from some appropriate sensors which are affected by motion attributes and characteristics of different body joints. Then, depending on these measurements, some features are extracted to be used in the process of training activity models, which in turn will be used to do inference later about these activities and performers. A typical example of a class of activities is those Activities of Daily Living (ADL) that people have the ability and custom for doing on a daily basis like eating, moving, hygiene, dressing, etc. [4].

There are many methods and data acquisition systems which are based on different sensory readings for recognizing these actions. Historically, heavy devices were used to collect accelerometer data such as data acquisition cards (DAQs) [5]. Later, smaller integrated circuits connected to PDAs were used for this aim, for instance, camera-based computer vision systems and inertial sensor-based systems. In computer vision, activities are recorded by cameras [6]. However, the drawback of the camera-based approach is that it may not work due to the absence of full camera coverage of all person's activities. In addition, for privacy issues, cameras are meddlesome as people do not feel free and comfortable being observed constantly.

Based on the *data acquisition* paradigm, HAR systems can be divided into two general categories: ambient/surrounding fixed-sensor systems and wearable mobile sensor systems. In the fixed setting, data are collected from distributed sensors attached to fixed locations in the activity's surrounding environment (where users' activities are monitored) such as surveillance cameras, microphones, motion sensors, etc; see [7] for applications in smart homes. Alternatively, the sensors are attached to interactive objects in order to detect the type of interaction with them, such as a motion sensors attached to the cupboard doors or microwave ovens (to detect opening/closing), or on water tap to feel turning it on or off, and so on. Although this method can detect some actions efficiently, it has many limitations due to its fixed nature. Another limitation is that if the user wants to leave the place, she will not be observed from the fixed sensors and her activities won't be detectable. Privacy is also another issue, especially when considering video surveillance cameras and/or auditory sensors.

Figure 1 shows the evolution of gathering inertial motion data; a National Instruments DAQ card-1200 with analog input, a USB-based wearable programmable chip, the more advanced devices, e.g., smart phones and smart watches.

In the *wearable*-based systems, the measurements are taken from mobile sensors mounted to (or embedded on commodity devices mounted to) human body parts like wrists, legs, waist, and chest. Many sensors can be used to measure selected features of human body motion, such as accelerometers, gyroscopes, compasses, and GPSs. They can also be used to measure phenomena around the user, such as barometers, magnetometers, light sensors, temperature sensors, humidity sensors, microphones, and mobile cameras. These kinds of sensors are essentially embedded on wearable devices. Consumer inertial MEMS IMU sensors are tiny devices with typical package size of 2.5 mm $\times$ 3 mm which contain three axial accelerometers and gyroscopes. They can be found in many commodity products we use almost every day as smartphones, fitness

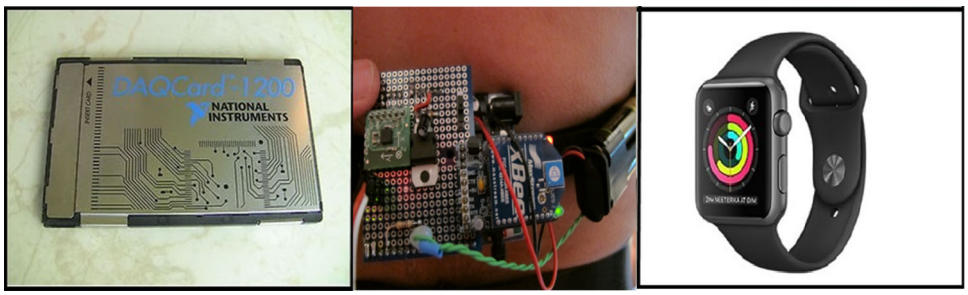


Fig. 1 Acceleration sensing hardware evolution

trackers, smartwatches, gaming controllers, etc. Two of those high tech IMU devices named BMI160 are inside the NASA helicopter Ingenuity which did fly many times on Mars surface in 2021.

On the contrary to the fixed-sensor-based systems, wearables are more intrinsic and able to continuously measure data from the user everywhere, while sleeping, working, or even traveling anywhere since it is not bounded by a specific place where the sensors are installed. Also, it is very easy to concentrate on *directly* measuring data of particular body parts efficiently without a lot of preprocessing that is needed, for example, in fixed depth cameras. So, as opposed to ambient sensors, a wearable sensor is intimately related to the body part to which it is mounted. So, in some sense it is directly expressive of this part (such as its motion dynamics, vital signs, etc.), without the need of much or any preprocessing at all. Examples of wearables include smartwatches, smart shoes, sensory gloves, hand straps, and clothing [8, 9]. Recent statistics show that the total number of smartphone subscribers reached about 6.5 billion in 2022 and is expected to reach 7.7 billion by 2022 [10]. According to an IDTechEx report [11] the wearable sensor industry is worth $2.5 billion in 2020, having tripled in size since 2014, and is expected to reach more than $4.5 billion in 2024. Therefore, the disadvantages of wearable mobile sensors of being intrusive, uncomfortable, and annoying have vanished to a great extent, making this method of on-board sensing from smart mobile and wearable devices very suitable for HAR data acquisition.

As the current trends have been greatly shifted towards systems that are data oriented with statistical machine learning analysis and synthesis, feature extraction and representation learning play essential roles in the effectiveness and efficiency of HAR systems. Feature engineering in timeseries inertial data has proven as effective as automatic feature extraction/selection using representation learning through the deep learning paradigm. From one perspective this is due to the limited datasets available and the small sizes of these datasets. From another perspective inertial timeseries data is low on explicit semantics compared, for example, with visual, auditory, and textual data; so it is not very much amenable to the high level convolutional operation dominant in deep neural networks. However, the power of deep learning emerged exclusively in *transfer learning* which is of great benefit in deploying practical solutions based on wearable and mobiles devices.

Transfer learning is the transfer of knowledge gained in one domain and/or task to another domain and/or task, so we needn't start over on the latter. Transfer learning is essential in HAR systems, not only for saving computational resources [by avoiding end-to-end training on the new task(s)], but more importantly for handling the tremendous heterogeneities induced by the huge configuration space incurred by the data collection and streaming processes in IMU-based HAR systems. These include among others the following:

1. Hardware heterogeneity:
   - Different kinds of devices (with embedded IMUs) such as smartphones, smartwatches, wrist bands, specialized IMU sensors that can be placed on different body parts, etc.
   - Different brands of the same kind of device. For example, varieties of smartphones such as iPhone, Galaxy, Huawei, etc.
   - Different models of the same brand of the same device. For example, iPhone 10, iPhone 11, etc.
   - All the above lead to different characteristics of the IMU sensing technology, sampling rates, dynamic ranges, sampling rate instabilities, etc.
2. Software heterogeneity:
   - Different operating systems; for example, iOS vs. Android.
   - Different versions of the same operating system.
   - Different filters embedded in the IMU sensors or sensors logs.
3. Data collection heterogeneity:
   - Variations in location(s) of the device(s) mounted to the body part(s). Given a fixed type of device, say smartwatch, it can be worn on the left or right wrist. Smartphones have more wider range of

locations. These lead to variations in the signals corresponding to the same activity.
- All other factors fixed, gender and age greatly affect the signals of the same activity.

Figure 2 gives a general framework of modern approaches to HAR that are data-oriented. The first process in the pipeline is, of course, the data collection which can be done through ambient devices such as surveillance cameras, microphones, etc., or through wearable devices such as smartwatches, smart clothing, etc. The data then are cleaned, curated, denoised, etc. Then, data go through higher-level preprocessing steps including, for example, the separation of body and gravitational acceleration. Data are then segmented, for example, into time spans based on gait cycle, or transitions between different consecutive activities, etc.

The data go through a crucial step for feature extraction/selection to capture and produce the main characteristics and representations of the given data. This process can be done in three main ways: (1) purely classical hand-crafted features, (2) purely automatic feature generation through deep neural methods, and (3) a hybrid of the two, where a preliminary simple feature extraction/transformation step is performed (for example, transforming the original time-series into a new series induced by calculating the auto-correlation function of the original signal), followed by representation learning through deep neural methods. The produced feature vectors/feature maps are then fed to a learning model which can be classical such as support vector machines or more recent deep neural networks. In the latter the last layer(s) are concerned with the application domain which can greatly vary between recognizing the type of the particular action performed, predicting and estimating the biometrics of the person performing the action such as the gender, age, weight, etc., the type of surrounding environment such as the type of ground on which walking occurs such as asphalt, grass, sand, etc.

To the best of our knowledge, we present the most recent research works related to the use of transfer learning in IMU-based HAR systems. The overall aim of this paper is to give an introduction and survey of human activity recognition systems using inertial motion data streamed from IMU sensors embedded on-board smart mobile and wearable devices. The survey is more oriented towards technical aspects with some hints to application areas. More specifically, we focus on the following:

1. We investigate the data collection process for HAR systems. In addition, we provide an elaborate detailed survey of all publicly available benchmark HAR datasets. To the best of our knowledge, this is the most comprehensive compiled list of available HAR datasets done so far.
2. A major preprocessing step for any HAR system is how to present the raw inertial data to the subsequent machinery of either a machine learning model or more traditional deterministic non-adaptive systems. Here we survey traditional handcrafted feature engineering techniques (extraction/selection) in addition to the more recent representation learning based on deep

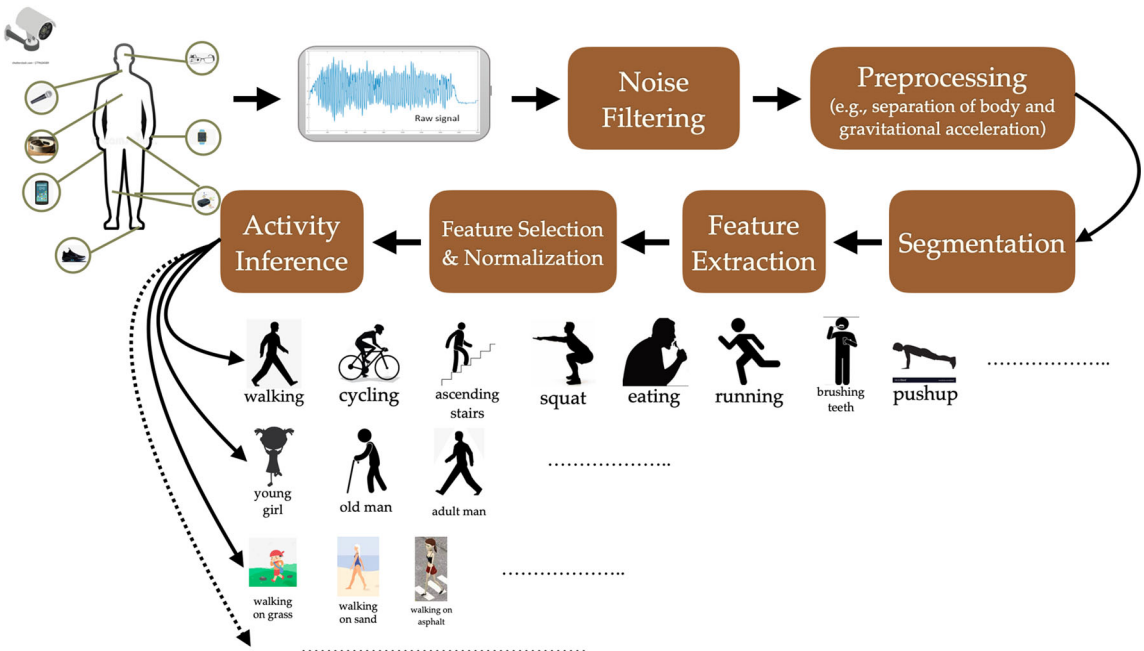


Fig. 2 General framework of HAR systems

neural computing. It is apparent that for the sole purpose of end-to-end classification models, more traditional techniques based on feature engineering are effective enough, however, for transfer learning for the purposes of cross-domain and/or cross-datasets and/or cross-environmental configurations and setups, deep learning is a necessity.

3. Transfer learning is a very powerful tool in machine learning in general (and a major path to artificial general intelligence), and in particular, for inertial-based HAR systems. This is due to the tremendous heterogeneity and the inherent gap between the data collection and streaming process during the system training and the corresponding process during deployment. This is further complicated by the low semantical content of timeseries inertial motion data, as compared, for example, with visual data. Here, we survey the work done in cross-domain HAR systems and how effective they are.
4. The following subject we address is based on a categorization of HAR systems depending on whether they are developed for online real-time prediction or for offline batch analysis.
5. Deep neural networks are highly parameterized complex models and they can be highly sensitive to small perturbations (chaotic systems), caused by, what is called, adversarial attacks. ("Chaotic" in the sense that the process of training deep neural networks can be viewed as the evolution of a dynamical system. This dynamical system is not robust, that is, it is sensitive to small perturbations. For example, small perturbations in the weights initialization, the round-off errors caused by floating point operations can lead to dramatically different results such as divergence of the training process instead of convergence, or convergence to better or worse solutions over the loss function, etc.) There are hardly few works addressing adversarial attacks against timeseries data. The hardest challenge is to qualify and quantify such attacks in this context. In visual data, adversarial attacks can be mainly characterized by the perturbed image being imperceptible to the human eye, a quality basically stemmed from the high semantical content of visual data. Such quality is absent from timeseries data. Here, we survey the few works addressing adversarial attacks against timeseries signals and further point to research challenges in this area.
6. We briefly discuss embedded implementation of HAR systems.
7. Finally, we discuss and survey some major applications domains of HAR systems based on inertial motion data.

As the HAR research area is very huge, we had to focus on particular aspects and give our own perspectives on what is going on. Of course, it is by no means comprehensive. However, it gives, at least, useful entry points for deeper investigation into the particular aspect of interest for the reader, in addition, to some tutorial and pedagogical merits. All the materials are presented with rather assessment and critical eyes.

The paper is organized as follows. Section 1 is an introduction. Section 2 discusses the data collection aspects and processes for human activity recognition tasks based on the inertial motion modality. In addition, we present, in a tabular form, a comprehensive survey of publicly available benchmark datasets in the literature related to that domain. This table is listed in "Appendix A". Section 3 introduces the feature extraction and selection techniques for timeseries data that are used in learning systems for activity analysis and inference. In Sect. 4 we discuss offline systems that are based on one-shot task, such as the recognition of a single isolated action, versus real-time continuous operation over the online streaming timeseries data. Section 5 surveys and assesses the state-of-the-art techniques for transfer learning of activity tasks based on inertial motion signals. Section 6 gives some typical application domains of activity recognition from wearable IMUs; this spans biometrics (e.g., age, gender, gait), virtual coaching, healthcare, and ergonomics. In Sect. 7, we discuss some implementations of HAR systems on embedded devices. Finally, Sect. 8 concludes the paper with an overall analysis of inertial-based HAR systems and their future potential.

2 Data collection

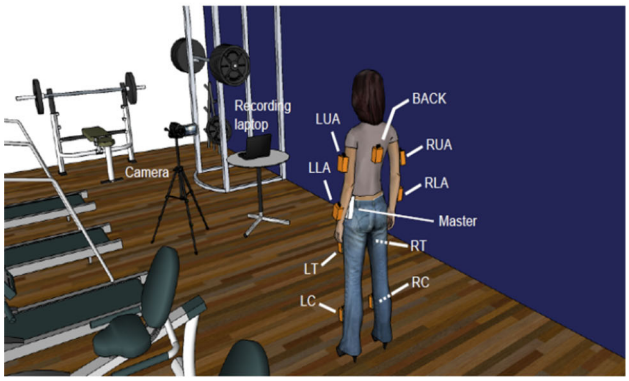
There is a wide variety of activity datasets that are based on inertial motion data. These are typically used as benchmarks to test new techniques for activity analysis and inference systems. A comprehensive survey and a rather elaborate description of publicly available benchmark datasets are given in Table 15 in "Appendix A". Figure 3 shows some configurations of different HAR datasets; (a) a motion-node that recorded data in USC-HAD dataset [12], (b) the sensor placement used for the REALDISP dataset [13].

Figure 4 shows sample images while performing different ADL activities (EJUST-ADL-2 [14]).

Figure 5 shows the activities and data collection of EJUST-ADL-1 dataset [24] using Apple smartwatch.

At designing any sensor-based activity recognition system, the number of sensors and their locations are very critical parameters. Regarding the locations, different body parts have been chosen from feet to chest. The selected locations are chosen according to the relevant activities. For

(a) A motion-node that recorded data in USC-HAD dataset [12].



(b) the sensor placement used for the REALDISP dataset [13].

Fig. 3 Configurations of different HAR datasets. Permission is taken from the authors at [12] and [13] to use this figure

example, ambulation activities (such as walking, jumping, running, etc.) are typically detected using waist or a chest sensor [15]. On the other hand, non-ambulation activities (such as brushing teeth, combing hair, eating, etc.) can be naturally classified more effectively using a wrist-worn, or in general upper body, mounted sensors [16].

Many kinds of *physical* features can be measured using wearable sensors [17]. These include: (1) environmental data such as barometric pressure, temperature, humidity, light intensity, magnetic fields, and noise level, (2) vital health signs such as pulse, body temperature, blood pressure, blood sugar (glucose) and respiratory rate, (3) location data which are typically identified by longitudes and latitudes using GPS sensors, and (4) body limbs motion such as acceleration and rotation of body parts like arms, legs, and torso using accelerometers and gyroscopes. Note that all of these can be considered as timeseries data. There are other different kinds of data that can be produced by wearable devices including textual data from social media.

However, in this article we focus exclusively on timeseries motion data. For the purpose of activity recognition, the inertial motion measurements have proven to represent human motion accurately, where some methods depending only on acceleration measurements have achieved very high accuracy [15, 16]. However, it is very difficult to extract the pattern of motion from acceleration raw data directly because of the high frequency oscillation and noise. So, preprocessing in addition to feature extraction and selection techniques should be applied on the raw data before modeling, or alternatively use deep learning methods for automatic feature extraction and selection (representation learning). The latter might not be a good choice over raw noisy data taking into consideration adversarial attacks, their qualitative and quantitative nature over timeseries data.

A line of work aims at automatically determining the on-body sensor location given the inertial motion signal of a given fixed activity. For example, Madcor et al. [18] study the particular action of 'walking' and how the induced signal can be used to determine the sensor's location. They do not do any feature extraction, instead, they use the raw signal data as a direct input to a random forest classifier. They validate their work on four datasets: EJUST-GINR-1 [19, 20], RealDisp [13, 21], HuGaDB [22], and MMUSID [23]. All of these datasets are either exclusively gait or including gait among other activities; however, only the gait is used in the study.

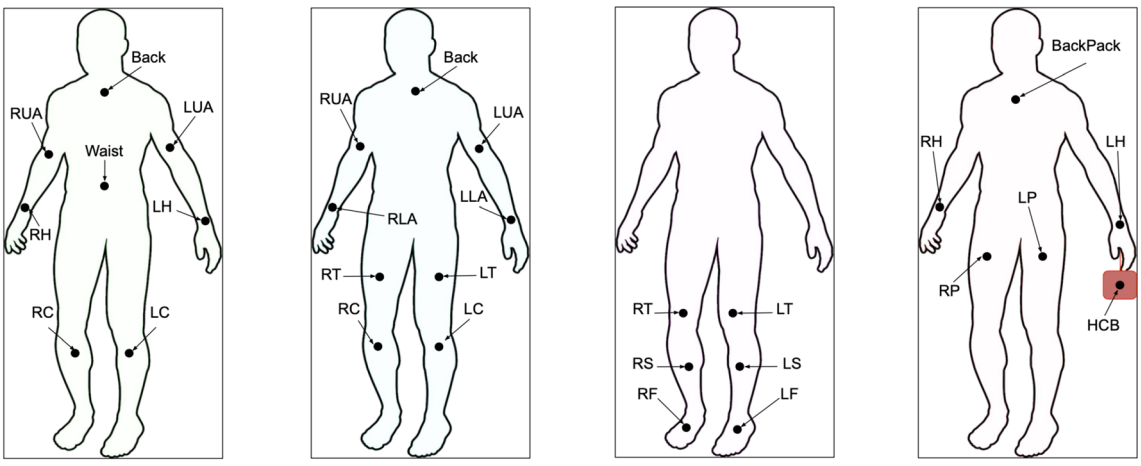
Figure 6 shows a silhouette diagram with sensors' locations in all the considered datasets. The signals used for recognition are the tri-axial linear acceleration in addition to the tri-axial angular velocity. These two

Fig. 4 Sample images while performing different ADL activities (EJUST-ADL-2 [14])

Fig. 5 The activities and data collection of EJUST-ADL-1 dataset using Apple smartwatch

modalities have proven effective enough to detect the sensor's location as evidenced by the performance metrics. Accuracy reached above 90% and most of the recall and $F1$-score results are above 80%. As expected most of the confusion in location determination occurs between symmetric sensors across the body such as between left and right calf, left and right wrist, etc. Such line of research can be applied in data augmentation of timeseries inertial data as well as in developing healthcare applications, for example, in treatment assessment for people with motion disabilities. Figure 7 shows a data sample of 3D acceleration, 3D rotation, and 3D angular velocity signals of the activity 'walk' (dataset EJUST-ADL-1 [24]).

A similar line of research is conducted by Abdu-Aguye in [25]. The output of this work can be used to create 'virtual IMU sensors'. In [25] the authors explore the feasibility of transforming/mapping inertial timeseries data streamed from a 'source' body location, whilst performing a given activity, to a different 'target' body location. They use a bidirectional long-short-term memory (Bi-LSTM) to learn such mapping. An LSTM is a rather stable form of recurrent neural networks (RNNs) that can capture long-term correlation between the current time sample with historical samples of the input timeseries. The bidirectional version takes as well the future into account when predicting the current moment, hence, it smooths the predicted



(a) EJUST-GINR-1 (b) RealDisp (c) HuGaDB (d) MMUSID

Fig. 6 Schematic diagram of on-body sensors locations in some activity datasets

signal, and hence, generates more accurate timeseries. However, apparently it can only be used in offline mode (or online mode with limited future input). The model is rather simple and is shown in Fig. 8. Figure 9 shows the IMU sensors placements on the human body (EJUST-GINR-1 dataset [19, 20]).

One of the most difficult and challenging problems in the collection of inertial motion data is synchronization. One reason that this process is more difficult in such data modality is that timeseries data has much lower semantics content, than, for example, visual or auditory data. In the latter modalities synchronization and/or segmentation can be done offline in a manual fashion, though it is a laborious process. Such offline manual process is not feasible in the case of timeseries data. Synchronization can involve determining the start and end of unit actions, the transition/switching between consecutive actions, synchronization and alignment among different IMUs used simultaneously in data collection, and in the case of multiple modalities the alignment among such modalities.

Fayez et al. [26] collect the VaIS dataset. It is a bimodal dataset of squats: visual and inertial motion timeseries. The authors use 9 different IMU devices mounted on different parts of the subject's body in addition to 2 RGB cameras one positioned in front of the subject and the other at half frontal/lateral. For the IMU devices that data are collected from the devices and then sent in the form of CSV files to a mobile app on an iPhone. Through this applications all of the IMU sensors are readily synchronized. Each record in the CSV files has a timestamp with a resolution in the order of milliseconds. Hence, this is used to synchronize among the 9 IMU devices. For the visual data they used the camera of a smartphone. They used a timestamp app in order to annotate the video capturing with a time format that is closest to that produced by the IMU devices. The readings obtained from the IMU units have the global timing in the following format <YYYY-MMM-DDTHH:MM:SS.S>.

Figure 12 shows a snapshot the video capturing with time annotation. Optical character recognition is then used on that section of the frames depicting the time annotation in order to extract the time. The time formats of both the

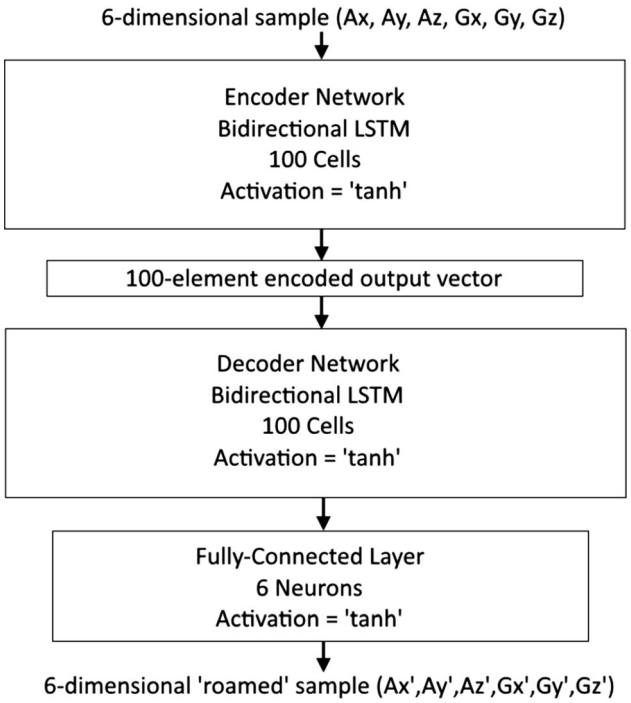
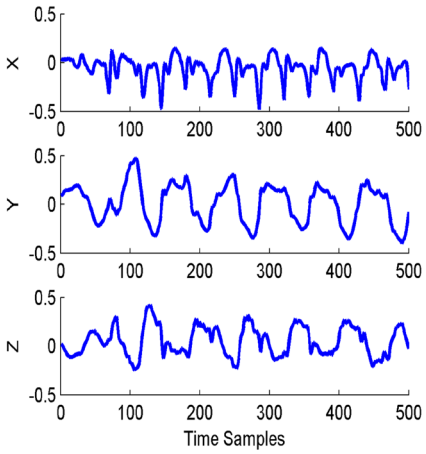
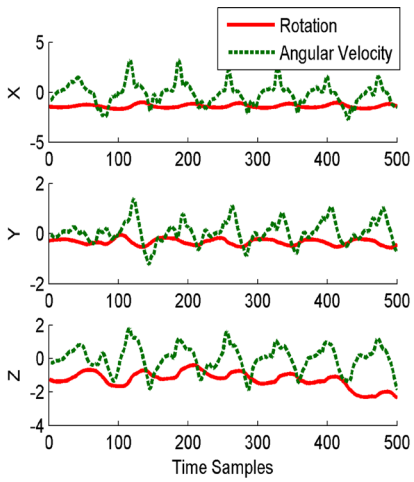


Fig. 8 Model for mapping inertial motion data across different body locations [25]

Fig. 7 A data sample of 3D acceleration, 3D rotation, and 3D angular velocity signals of the activity 'walk' (dataset EJUST-ADL-1 [24])



(a) A data sample of 3D acceleration (g force) signals of the activity 'walk'.



(b) A data sample of 3D rotation (rad) and angular velocity (rad/s) signals of the activity 'walk'.

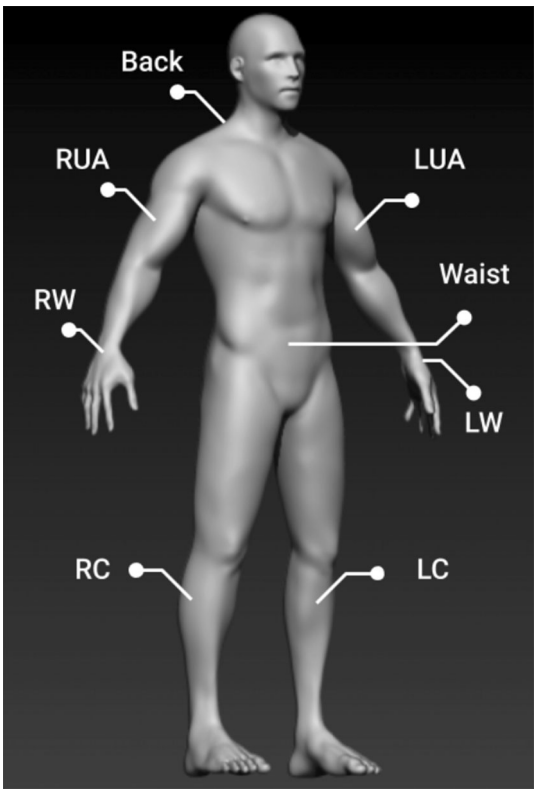


Fig. 9 IMU sensors placements on the human body (EJUST-GINR-1 dataset [19, 20])

IMUs and that extracted from the frames are matched together. They finally matched the time instances in both modalities taking into consideration the difference in IMUs sampling rate and the video streaming frame rate. This caused a delay error between the two modalities that does not exceed 200 ms. During the data collection process, the authors used a whistle to indicate exactly when the subject started the activity. This would generate a high frequency peak in the sound wave used to cut the video indicating the execution of the given activity. The first extracted frame is then matched with the corresponding IMUs frame given the aligned time steps. Similar procedure employed to indicate the last video and IMUs frames.

The same authors have recently collected a new workout activity dataset [27]. The dataset is called *MMDOS*; it is a multi-modal dataset and is unique in the sense that it combines many data modalities and sensors including: RGB, inertial motion, depth, and thermal. Accordingly, the synchronization here is much more challenging than in the previous work. The data are collected from 50 participants performing four types of workout exercises. As in the previous work all the IMU sensors are already synchronized through a mobile app running on an iPhone. All IMUs have the same sampling rate. The depth imaging is obtained using a Kinect camera, and the timestamp is extracted from the Kinect metadata using MATLAB with the same format as the IMUs. RGB video are captured using mobile phones and a mobile app is used to annotate each frame of the video capture with the timestamp.

See Fig. 10 for an illustration of time annotation of RGB video frames. Optical character recognition is used to extract this annotation and convert into proper time format. The timestamp of the frontal camera is used as a frame of reference for all other devices, taking into account the different sampling rates of the involved sensors. In addition, the authors used a beep sound to indicate exactly when the subject starts the activity so that it makes a peak in the audio signal. This sound is used to align the start of the RGB video, and all timestamps are then synchronized taking into consideration the different sampling rates across all devices. As for the thermal imaging they performed the synchronization completely manually due to the inability to receive any metadata, including timestamp, from the video streaming. The whole synchronization process is depicted in Fig. 11. Figure 12 shows a video recording with timestamp annotation [26].

3 Feature extraction/selection

For inertial-based HAR systems, many approaches to feature extraction and selection have been developed focusing on the structural and/or statistical properties of the input timeseries. In this section, we survey the main feature extraction and selection techniques. There are different kinds of features used to represent timeseries data. Some typical examples of features are shown in Table 1, which are essentially based on handling the timeseries either in the time domain or in the frequency domain.

Huynh et al. [29] have suggested that selecting individual tailored features for each activity can improve both the computational efficiency and the recognition accuracy. But, designing a specific descriptor for each activity is not flexible

Fig. 10 Frontal camera view of a shoulder press exercise with the timestamp on the lower right [27]

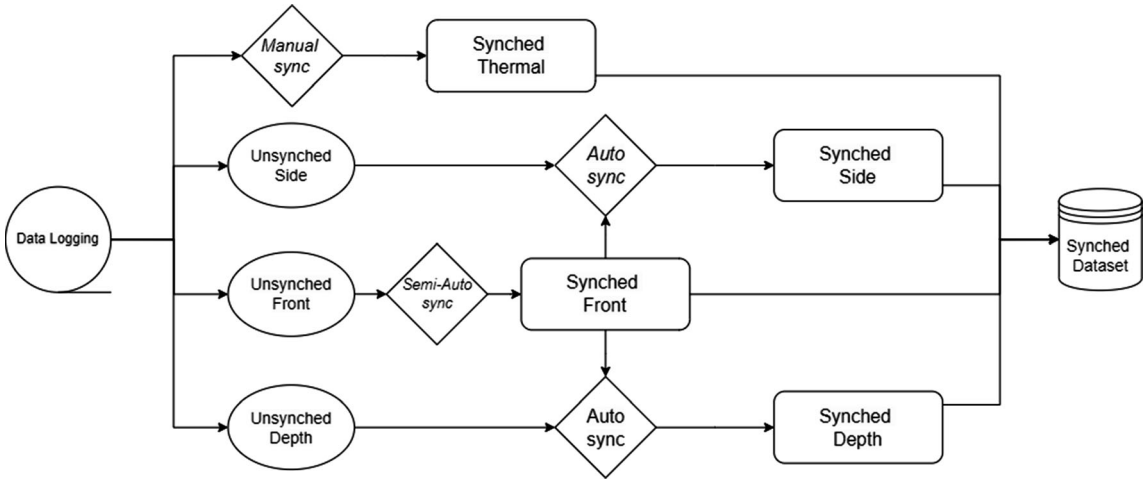


Fig. 11 Synchronization among all modalities in the MM-DOS dataset [27]

for adding new activities to the system. For example, if the system has fifty activities, fifty different descriptors are needed to design in the worst case, which can be impractical.

In the following sub-sections, we survey research works using different types of features, e.g., autocorrelation function, Fourier analysis/spectra, autoregression, raw data, embedding, etc. These types of features can be used standalone or combined with other types of features, e.g., time domain features (standard deviation, variance, kurtosis, root mean square, mean, skewness, etc.) and/or frequency domain features (discrete cosine transforms, wavelet transform, entropy, quantiles, etc.).

Fig. 12 Video recording with timestamp annotation [26]

3.1 Autocorrelation function

One of the most effective and simple feature extraction techniques is based on the use of autocorrelation function [30] taking lag values starting from 0 up to a certain maximum value. The autocorrelation function computes the correlation/similarity between a given timeseries and a lagged version of it. The empirical estimation of the autocorrelation function can be computed as follows:

$$\hat{\text{acf}}(\ell) = \frac{\hat{\gamma}(\ell)}{\hat{\gamma}(0)}$$
$$\hat{\gamma}(\ell) = \frac{1}{T}\sum_{t=1}^{T-\ell}(x_{t+\ell} - \bar{x})(x_t - \bar{x}) \quad (1)$$

where $\bar{x}$ is the sample mean of the timeseries and ℓ is the lag value (the amount by which the original signal is shifted). The autocorrelation function captures the linear correlation between the timeseries and the shifted versions of it. In some way it can be viewed as a feature reduction technique that reduces the original signal into a smaller signal of length depending on the maximum lag value ℓ. From another perspective, it can be as well considered as length normalization, that is, different timeseries may have different lengths, however, if we fixed a certain lag, then the acf of all of them up to this lag will produce timeseries with the same length. This might be necessary for many analytical techniques, such as when required to find the similarity/dissimilarity between two timeseries.

Table 1 Different types of timeseries features

Technique	Feature
Time domain features	Standard deviation, variance, kurtosis, root mean square, mean, skewness, total energy, range (maximum–minimum) [28]
Frequency domain features	Fourier transform, discrete cosine transforms, wavelet transform, entropy, quantile 0.25, quantile 0.50, quantile 0.75 [16]

For example, Gomaa et al. [24] used a smartwatch to collect sensory data for 14 activities of daily living (ADLs), creating a dataset called EJUST-ADL-1 (see Table 15 for full description). They compute the autocorrelation function for each of the tri-axial timeseries of the acceleration, angular velocity, and rotation displacement, up to a certain lag (empirically choosing 10 as suitable). These are vectorized and fed to a random forest (RF) classifier that recognizes the performed activity. They achieved 80% accuracy over the 14 activities.

The autocorrelation function has also been effectively used for identifying periodic peaks [31] through segmenting repetitive actions. In a different context Mostafa et al. [32] extracted the autocorrelation function from a tri-axial accelerometer (linear motion) and a tri-axial gyroscope (angular motion) to produce a multi-channel feature vector consisting of 6 channels each of length 11 (10 is the maximum lag of the acf). The signals are streamed from IMU sensors during walking activity. The authors experimented with multiple gait datasets including: EJUST-GINR-1 [19, 20], OU-ISIR [33], Gait Events DataSet (GEDS) [34], and Human Gait Database (HuGaDB) [22].

The acf features are then fed into two models for biometrics predictions: a classical random forest for gender classification and a convolutional neural networks (CNN) for age estimation. Figure 13 shows the CNN architecture used. In comparison with other works in the literature, their CNN regressor performs better over almost all the datasets considered with age mean-absolute-error ranging from 0.9 years up to 2.4 years. For gender recognition both the random forest and CNN have comparable results above 90% (except for the OU-ISIR dataset). This indicates the effectiveness of the autocorrelation function as timeseries representation which is rather simple to compute and interpret.

Adel et al. [20] use the autocorrelation function as a feature extractor for person identification using inertial streaming from wearable IMUs during walking activity. Again the acf is computed for the 6-axial inertial signal (3-axial accelerometer and 3-axial gyroscope). They experiment with the EJUST-GINR-1 gait dataset [19, 20]. It is apparent from Fig. 14 the inherent periodicity of the gait signal, specially in the y-direction, and hence, the acf is well suited to capture the gait characteristics of the given timeseries, specially capturing the corresponding unique gait fingerprint of each individual. The authors feed the induced feature vectors to three classifiers: a SVM, a RF, and a CNN. The number of subjects are rather limited (20), however, the authors study the best on-body sensor location for person identification and how fusion of several IMU sensors mounted on different body parts would affect the results. Even though the case for gait is obvious (lower body locations), the proposed methodology open new directions for similar kinds of research with more complex types of activities.

A hidden assumption in the use of autocorrelation is that the activity signals are statistically *stationary*. It means that the statistical behavior of the timeseries does not change over time. To the best of our knowledge this has empirically shown to be the case where the performance of predictive models is rather satisfactory. However, much work still needs to be done to verify that assumption and for which classes of activities. And more importantly to explicitly state this assumption and put it in perspective. Another assumption, or maybe rather simplification, is that each inertial modality (whether linear or angular), and each axis of the same modality is assumed to be completely independent of the others. Maybe a better descriptor would be obtained by considering the cross-correlation among the different axes of the same modality and even across different modalities.

Nourani [35] presented a comprehensive comparison of HAR activities using IMU sensors; Nourani selected five state-of-the-art feature-sets as case studies; statistical features, self-similar features, histogram bins features, physical features, orientation invariant features. The self-similar features includes the number of autocorrelation peaks, prominent autocorrelation peaks, weak autocorrelation peaks, maximum autocorrelation value, first autocorrelation peak height, etc.

Wang [36] presented a data fusion-based hybrid sensory system for elderly ADL recognition; one of the proposed features is the autocorrelation of the data collected from the accelerometer, gyroscope, magnetometer, barometer, and temperature sensors. Janidarmian et al. [37] presented a comprehensive analysis on wearable acceleration sensors in HAR; one of the features used is the autocorrelation; the height of the first and second peaks and the position of the second peak.

Morris et al. [38] proposed RecoFit a system that uses a wearable sensor to find, recognize, and count repetitive exercises. Data was acquired by a wearable armband worn on the subject right forearm; it contains a SparkFun IMU sensor (3-axial accelerometer and 3-axial gyroscope). The authors used the autocorrelation function to find patterns of self-similar repetitive exercise. Each 5-s window is transformed into 224 features. For instance, the authors compute the following five features; number of autocorrelation peaks, prominent peaks, weak peaks, maximum autocorrelation value, height of the first autocorrelation peak after a zero-crossing.

Muehlbauer et al. [39] proposed a HAR system based on segmentation, recognition, and counting. They achieved 85% segmentation accuracy and 94% recognition accuracy on ten classes based on user-independent training. The proposed approach for segmentation utilizes features based

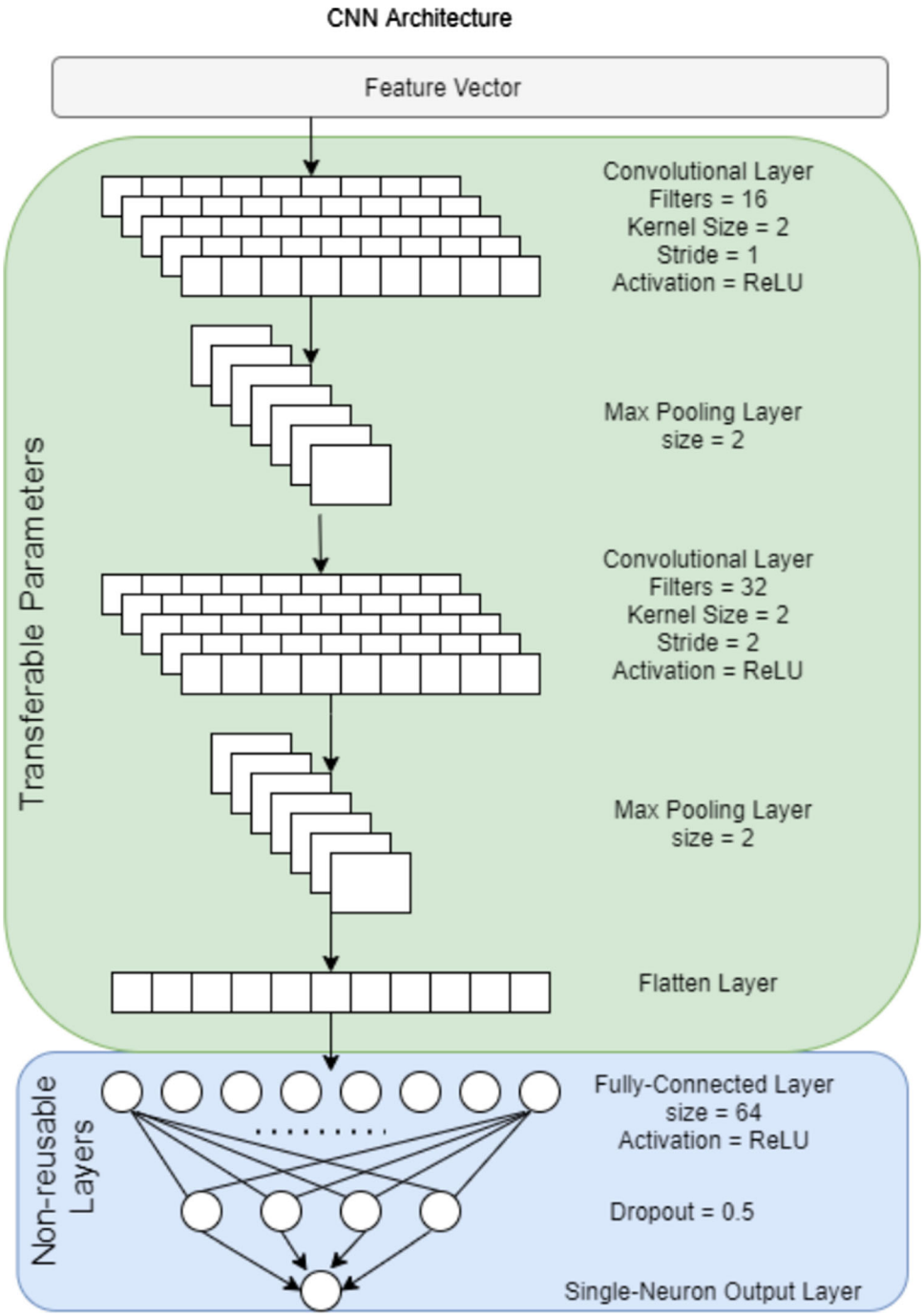


Fig. 13 The BioDeep CNN architecture [32]

on an autocorrelation function. For each window, the authors calculated the median crossings and peaks of the signal autocorrelation function. The high values of these two features indicated clearly that an exercise was performed.

Zebin et al. [40] proposed a HAR system based on multichannel time series signals collected by five wearable sensor units each with an integrated MPU-9150 IMU (accelerometer and gyroscope). Six activities were recognized; walking, walking upstairs, walking downstairs, sitting, standing, lying down. The authors applied four machine learning algorithms; k-NN, NB, SVM, and MLP. Several feature extraction methods were adopted; average value for all three acceleration components, RMS value for all three acceleration components, autocorrelation features (height of main peak, height and position of second peak), spectral peak features (height and position of first six peaks), spectral power features (total power in five adjacent and pre-defined frequency bands).

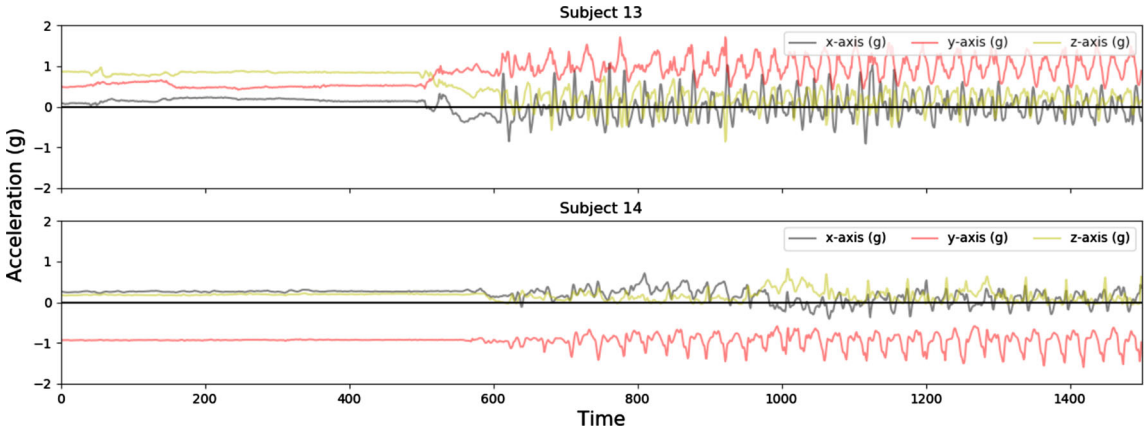


Fig. 14 A sample of gait signal (EJUST-GINR-1 gait dataset [19, 20])

Jain et al. [41] proposed a stride segmentation method from 3-axial accelerometer data for three walking activities (level walking, walking upstairs, and walking downstairs). The autocorrelation function is used for stride boundaries estimation. The segmentation and extraction of stride relevant data were achieved using a tuning parameter based on the minimum standard deviation among the segmented strides. The method evaluation is performed on the HAPT [42] and OU-ISIR gait [43] inertial sensor datasets.

Bota et al. [44] proposed a Semi-Supervised Active Learning (SSAL) based on Self-Training (ST) approach for HAR to automate partially the annotation process in order to reduce the required number of annotated samples by an expert. The proposed approach selects the most relevant samples for annotation and propagate their annotation to the most confident samples. The authors compared supervised and unsupervised models with the proposed approach on two ADL datasets; (1) the UCI dataset [45] with the following activities; standing, sitting, downstairs, upstairs. The considered features are statistical domain features (e.g., skewness, max, mean, histogram), temporal domain features (e.g., zero crossing rate), spectral domain features (e.g., median frequency, max power spectrum). (2) the CADL dataset (collected by the authors) with the following activities; standing, sitting, downstairs, upstairs. The considered features are statistical domain features (e.g., mean, histogram, root mean square), temporal domain features (e.g., autocorrelation, temp centroid), spectral domain features (e.g., max power spectrum). The proposed method reduced the required annotated samples by more than 89% while maintaining an accurate model.

Altun and Barshan [46] compared various techniques for HAR using inertial and magnetic sensors; Bayesian decision making (BDM), least-squares method (LSM), k-NN, DTW, SVM, ANN. Daily and sports activities are recognized using five sensor units. The sensor units are worn by eight users on chest, arms, and legs. Each sensor unit comprises 3-axial gyroscope, 3-axial accelerometer, 3-axial magnetometer; 26 features were calculated for each signal: minimum and maximum values, mean value, variance, skewness, kurtosis, 10 equally spaced samples from the autocorrelation sequence, first five peaks of the discrete Fourier transform and the corresponding frequencies.

Yu [47] presented CNN and RNN models to classify fall and non-fall activities. A publicly available dataset was used [48]; 6 MTx sensor units were worn; each sensor unit comprises 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer and atmosphere pressure; this dataset includes 20 fall activities and 16 ADL activities. CNN achieved an accuracy of over 95% for classifying fall and non-fall activities. RNN achieved an accuracy of over 97% in classifying fall, non-fall, and a third category (defined as "pre/postcondition"). The following features were extracted for each time window: minimum and maximum values, mean value, variance, skewness, kurtosis, first 11 values of autocorrelation sequence, and the first five peaks of discrete Fourier transform.

Pereira et al. [49] presented a biofeedback system based on sEMG and IMU sensors for home-based physiotherapy exercises evaluation. sEMG signal was used to identify temporal intervals in which muscular activity was occurring. Exercise repetitions were segmented into time windows where features related to subject posture were extracted. The extracted features include; statistical features (e.g., skewness, kurtosis, histogram, etc.), temporal features (e.g., mean, median, maximum, minimum, variance, temporal centroid, standard deviation, root mean square, autocorrelation, etc.). These features were input to DT, k-NN, RF, and SVM methods to classify proper/improper exercise execution and achieved an accuracy of $\geq 92\%$.

Rosati et al. [50] compared two feature sets for HAR; the first includes time, frequency, time-frequency domain features, while the second feature set includes time domain

features considering the signals physical context. The comparison of the 2 feature sets were based on 4 machine learning classifiers performance. 61 healthy subjects performed 7 ADL (activities of daily living) activities while wearing an IMU-based device. A 5 s window was used to segment each signal; for each window 222 and 221 features were extracted for each feature set, respectively. The first feature set includes 222 features from different domains; 20 time-domain features (mean value, variance, standard deviation, skewness, kurtosis, minimum value, maximum value, 25th percentile, 75th percentile, interquartile range, 10 samples from the autocorrelation sequence), 3 frequency-domain features (mean and median frequency of the power spectrum, Shannon spectral entropy), 14 time-frequency domain features (norms of approximation and detail coefficients of 7 levels of decomposition of discrete wavelet transform). Each feature set had simultaneous feature selection and classifier parameter optimization performed using a Genetic Algorithm (GA). The results showed that the SVM has the best performance on both feature sets (97.1% and 96.7%). The second feature set allowed for easier understanding of changes in the biomechanical behavior in complex cases, e.g., the application on pathological subjects.

Muaaz [51] proposed a HAR system called WiWeHAR using Wi-Fi and wearable IMU sensors. The standard Wi-Fi network interface cards collected the Channel State Information (CSI) and the wearable IMU (accelerometer, gyroscope, magnetometer) collected the subject's local body movements. The authors computed the time-variant Mean Doppler Shift (MDS) from the processed CSI data and the magnitude from each IMU sensor. Then, they extracted 23 time-domain and frequency-domain features from the MDS and the magnitude data: mean, variance, standard deviation, skewness, mean absolute value, waveform length, enhanced mean absolute value, enhanced waveform length, weighted mean absolute value (window function step-wise or continuous), maximum fractal length, mean amplitude change, root mean square, difference absolute standard deviation, simple squared integral, Willison amplitude, zero crossing, slope sign change, max of the absolute value, slope, peaks of the autocorrelation, spectral peaks, spectral energy. The authors evaluated the performance of WiWeHAR by using a multimodal human activity dataset that was collected by 9 subjects. Every subject performed four activities: walking, falling, sitting, picking-up object from floor. The results indicated that the proposed multimodal system outperformed the unimodal individual CSI, accelerometer, gyroscope, and magnetometer HAR systems achieving overall recognition accuracy of 99.6–100%.

3.2 Fourier analysis

Another major, and very classical descriptor, of the timeseries is based on Fourier transform and its variants. Generally, Fourier transform can be considered as the projections of the given timeseries over an infinite basis of complex exponentials $e^{i2\pi nt}$. As these complex exponentials are orthonormal they can be considered to form a basis of an infinite dimensional space. So, the Fourier coefficient is a projection. In other words the nth Fourier coefficient is a projection of the function $f(t)$ over the axis of nth complex exponential $e^{i2\pi nt}$. The numerical/computational realization of the Fourier transform is the Discrete Fourier transform, and its most prominent implementation is called the Fast Fourier transform. The coefficients of the transform can be taken as a descriptor/feature vector of the input timeseries. As the transform is infinite in nature, we consider only the dominant components which correspond to the dominant frequencies (in magnitude) in the original timeseries.

Of particular interest is the *wavelet transform*. Wavelet transform is an improvement over the classical Fourier transform where the signal now can be decomposed in terms of both time and frequency localized basis functions. This helps capture the irregularities and non-stationarities in the input timeseries as it is a localized Fourier transform. Given a continuous-time signal $x(t)$, its idealized Fourier transform can be computed as follows:

$$\gamma(s,\tau) = \frac{1}{\sqrt{s}} \int x(t) \cdot \psi\left(\frac{t-\tau}{s}\right) \mathrm{d}t \tag{2}$$

The function ψ represents the *wavelet basis function* (called the Mother wavelet). Its wavelet is characterized by two parameters: its translation value τ and its scaling value s. These parameters determine both the time-localization and the frequency-localization of the mother wavelet. By continually changing the parameter τ, the mother wavelet is translated and accordingly temporal resolution is achieved. On the other hand, by continually changing the parameter s resolution in the frequency ranges is gained.

By convolving the given timeseries with the scaled and translated version of the mother wavelet, a similarity value between the signal and the mother wavelet at that translation and scale is obtained. This similarity defines the *wavelet coefficient* $\gamma(s,\tau)$. Abdu-Aguye et al. [52] did investigate the effectiveness of two feature extraction techniques. One of them is based on decomposing the signal applying wavelet transform using the Haar mother wavelet.

The autocorrelation of the resulting coefficients are computed up to a certain lag (they experimented with 5, 10, and 15). The resulting feature vectors are then fed to 4

different classifiers: random forests [53], canonical correlation forests (a variation of random forests) [54], ProtoNN [55], and Deep Forests (gcForest) [56].

The problem explored is direct activity recognition, specifically, recognizing activities of daily living over three datasets: HAPT [42], Sports and Daily Activities [57], and EJUST-ADL-1 [24]. Their results show the effectiveness of the developed descriptor based on the wavelet transform, specially with comparison with the other descriptor, proposed by the authors, based on autoregressive modeling. Across all classifiers and datasets the wavelet-based descriptor achieves at least in the range of 80%'s.

Motivated by the strong need of monitoring human activities for healthcare, Preece et al. [58] have done a rather extensive comparison of different types of feature descriptors extracted from accelerometer signals. They compare 14 feature extraction methods derived from wavelet transform, time-domain, and frequency-domain characteristics of the inertial signals. The comparison was done using two activity datasets collected from 20 subjects. One dataset contains the common daily activities of walking, ascending stairs, and descending stairs. The second dataset is more encompassing and contains 8 activities including hopping, running, jogging, and jumping. To test the sensitivity of the feature descriptors with respect to the sensor's location, the authors compared the activity classification accuracy over inertial signals streamed from three different lower limb locations: waist, thigh, and ankle, and combinations of these. The lower limb was chosen exclusively as all the activities mainly involve the lower body joints. Nearest neighbor classifier was used to test the performance of different configurations of feature descriptor/sensor placement/activity set. Cross-subject validation is used to assess and compare amongst the different configurations. Their empirical study indicated that the wavelet transform can best be used to characterize non-stationary inertial signals, however, it does not perform as well for classifying stationary activities performed by healthy subjects. In the latter case frequency-based features perform supreme with respect to others. Their best descriptor performed about 95% cross-subject classification accuracy.

Foerster and Fahrenberg [59] proposed motion and posture recognition model based on accelerometers data. The considered activities include posture, motion, stairs, lying, supine, sitting, walking, and bicycling. The activities were recorded by 31 subjects comprising 13 motions and postures that were repeated for three times. Five uni-axial sensors and three placement locations (sternum with three axes and right and left thighs). The authors proposed a hierarchical classification method that recognizes sub-categories of motion and posture activities with only 3.2% mis-classifications. The raw DC values and rectified AC values were averaged for each activity and monitoring segment. The walking frequency was calculated using short-time Fourier transform within a frequency band of 0.5–4 Hz from the sternum sensor z-axis.

Preece et al. [58] compared different feature descriptors extracted from accelerometer signals. They compared 14 feature extraction methods derived from wavelet transform, time-domain (e.g., mean, standard deviation, median, etc.), and frequency-domain (e.g., principal frequency, spectral energy, magnitude of first five components of FFT analysis, etc.) of the IMU signals. In order to extract the frequency-domain features, an FFT was applied on each 2-s window (128-sample) with 50% overlap between consecutive windows. The final frequency-domain feature set comprises of the magnitudes of the first five components of the FFT power spectrum.

Janidarmian et al. [37] proposed a HAR system based on wearable accelerometers. The proposed features include the Power Spectral Density (PSD), the Fourier coefficients in the frequency domain, the positions and power levels of the highest 6 peaks of PSD computed over a sliding window, and the total power in 5 adjacent and pre-defined frequency bands.

Bao and Intille [60] proposed a HAR system based on subject annotated acceleration data. Features like mean, energy, frequency-domain entropy, and correlation are extracted from 512 sample sliding windows of acceleration data with 50% overlapping between consecutive windows. Each window duration is 6.7 s. A several seconds window is sufficient to cover cycles in activities like walking, window scrubbing, and vacuuming. The window size of 512 samples allows for quick computation of FFTs.

Huynh and Schiele [29] studied the impact of features and window length on HAR accuracy. To study the effect of various windows lengths, acceleration features were extracted from windows of 128, 256, 512, 1024, and 2048 samples. For each window, the authors computed the magnitude of the mean, variance, energy, spectral entropy, and the discrete FFT coefficients. The results showed that there is neither a single feature nor a single window length that achieved best accuracy for all activities. The FFT features were always among the features with the highest cluster precision. However, the FFT coefficients that achieved the highest precision differ for each activity, and the recognition can be enhanced by selecting features for each activity individually. In general, the highest peaks for the FFT coefficients were laid between the first and the tenth coefficient. The recognition results indicated as well that integrating different FFT coefficients to bands of exponentially increasing size can be a trade-off for using individual or paired coefficients. For the non-FFT features, the variance has continuously high precision values, however, the spectral entropy has the highest values for the

activity 'standing'. On average, the features with window lengths of 1 and 2 s achieved slightly higher precision values than those with other window lengths. However, there are big differences across the various activities, and selecting different window lengths for various activities resulted in higher recognition rates.

Eyobu and Han [61] presented a spectrogram-based feature representation and data augmentation technique for HAR based on wearable IMU sensor data. The Hanning window was utilized and 512 samples were used in each FFT vector with a sampling frequency of 50 Hz. The Hanning window can be defined as in Eq. (3) and was named for Julius von Hann, an Austrian meteorologist. It is also known as the Cosine Bell.

$$w(n) = \cos^2\left(\frac{n}{N}\pi\right), \quad n = -\frac{N-1}{2}, \ldots, -1, 0, 1, \ldots, \frac{N-1}{2} \tag{3}$$

An ensemble of data augmentations in feature space is proposed to tackle the data scarcity problem. Performance evaluation was done on an LSTM architecture to assess the impact of the proposed feature representation and data augmentations on HAR accuracy. In addition to evaluating on a newly proposed dataset, the data augmentation technique is evaluated on the UCI HAR dataset [62]. Using few spectral features, the authors achieved state-of-the-art recognition performance. Fewer spectral features exhibit a lesser training time of about 1 h and 45 min compared to about 2 h and 30 min for the large feature set utilized in this work. The proposed extraction approach achieved the best performance enhancement in accuracy of 52.77% compared to the UCI online dataset. The proposed OR (Original spectral features) + LA1 (1st Local Averaging) achieved the best performance enhancement in accuracy of 32.6% compared to the UCI online dataset.

García [63] proposed human activity recognition and segmentation using Hidden Markov Models. The recognition is based on IMU signals collected by two smartphone sensors: accelerometer and gyroscope. Six different activities are considered: walking, walking-upstairs, walking-downstairs, sitting, standing, and lying. These activities were performed by 30 users. The feature vector included: mean, standard deviation, median, absolute value (for each axis), minimum and maximum values, Signal Magnitude Area (SMA), energy average sum of squares, interquartile range, entropy, autoregression coefficients, correlation coefficient between two signals, maximum frequency component, frequency signal weighted average, skewness, kurtosis, frequency interval energy within FFT 64 bins of each window, angle between two vectors. The author concluded that using Cepstral techniques were suitable for HAR and segmentation. Despite inertial analysis being similar with speech recognition, some transformations (like pre-emphasis) are useless for inertial signals analysis. The RASTA filtering and delta coefficients were found to be only suitable for activity segmentation. Thus, their use is not recommended when the subject/activity recognition is the main goal. The use of Activity Sequence Modelling is highly suggested since it reduced all error rates for recognition and segmentation. The subject recognition error rate was reduced from 26.8 to 16.5% and the vector size from 561 to 180 features. The activity segmentation error rate was reduced from 5.2 to 0.4% and the vector size from 561 to 405 features. The activity recognition error rate was increased slightly from 3.5 to 3.6% using the Cepstral techniques, however the vector size was lower (only 180 features).

3.2.1 Spectra

Machado et al. [64] proposed a HAR system based on a 3-axial accelerometer sensor; the proposed system is based on unsupervised learning approach. The considered features are statistical domain features (kurtosis, skewness, mean, standard deviation, interquartile range, histogram, root mean square, median absolute deviation), temporal domain features (zero crossing rate, pairwise correlation, autocorrelation), spectral domain features (maximum frequency, median frequency, cepstral coefficients, power spectrum, Mel-frequency cepstral coefficients, fundamental frequency, power bandwidth).

Krause et al. [65] proposed a wearable system to specify online the user context and context transition probabilities. The authors tried various sampling frequencies and different feature engineering methods were applied. For low sampling rates (1 and 10 samples per minute), the onboard averages and absolute differences were calculated. For high sampling rate (8 samples per second), the Fast Fourier Transform (FFT) oscillatory spectra of the recorded accelerometer values were calculated. A 64 point FFT (i.e., spectra of 8 s windows with different spectrum for each axis) is a good choice to discriminate various fine-grained movement patterns. The resulting spectra were log-transformed. In the low sampling rate cases, all 8 data channels of the BodyMedia device were recorded forming an 8-dim feature-space. For the high sampling rate, the spectra of the two accelerometers were recorded forming a 128-dim feature-space.

Abreu et al. [66] proposed a HAR system where the signals were collected from four sensors: accelerometer, gyroscope, magnetometer, and microphone. A new dataset was proposed comprising 10 activities; e.g., opening door, brushing teeth, typing on keyboard, etc. The proposed HAR classifier is based on multiple HMMs one per activity. The authors extracted all the three axes (x, y, z) of

the 3-axial sensors (accelerometer, gyroscope, and magnetometer) and the overall magnitude of each $\sqrt{x^2+y^2+z^2}$. To address similar activities classification conflict (e.g., opening door and opening faucet), the sound was considered in the classification process; the authors combined the IMU sensors with the microphone. Each signal is segmented in time windows of length 250 ms without overlap. For each 250 ms window, over 30 different features were extracted. The features came from temporal, spectral, and statistical domains that were calculated for all sensors and axes. The final output is a feature vector with 265 features for each 250 ms time window. The authors achieved an overall accuracy of $84 \pm 4.8\%$ utilizing 27 features selected by Forward Selection coming from different domains (statistical, temporal and spectral). In online recognition, the proposed solution experienced preliminary tests by 3 of the 8 initial subjects. The classifier detected activities within a continuous stream with an $F1$-score of $74 \pm 26\%$.

Shoaib et al. [67] present a fusion-based HAR system using smartphone motion sensors. Ten male subjects perform seven activities; walking, running, sitting, standing, jogging, biking, walking upstairs, walking downstairs. Five smartphones (Samsung Galaxy SII i9100) are used with the following sensors: accelerometer (with and without gravity), gyroscope, magnetometer. The authors adopted various machine learning techniques: Bayesian networks, SVM, logistic regression, k-NN, decision trees, and random forest. Three feature sets were proposed of time domain features and one set of frequency domain features (the sum of the first 5 FFT coefficients and spectral energy). The authors showed that the utilized sensors, except the magnetometer, were individually able of recognition based on the activity type, the body locations, the feature set, and the classification method (personalized or generalized). The authors also showed that the sensors fusion only enhances the overall recognition accuracy when their individual performances are not high enough.

Tahir et al. [68] presented a HAR model based on IMU sensors, i.e., gyroscopes and accelerometers. The IMU data is processed using multiple filters such as Savitzky–Golay, median, and Hampel filters for examining the lower/upper cutoff frequency patterns. Statistical features (average, middle, squared deviation, and max/min values), frequency features (chirp z-transform (CZT), spectral entropy, the Helbert transform), and wavelet transform features (the Walsh–Hadamard transform) were extracted. The Adam [69] and AdaDelta [70] methods were utilized in the feature optimization phase. The proposed model was evaluated on the USC-HAD dataset [12] and the Intelligent Media-sporting Behavior (IMSB) dataset [71] that is a new subject-annotated sports dataset. The "leave-one-out" cross validation is utilized and the proposed model achieved recognition accuracy of 91.25%, 93.66% and 90.91% when compared with the USC-HAD [12], IMSB [71], and Mhealth [72] datasets, respectively.

Figueira et al. [73] proposed a HAR system where two sensors were used (accelerometer and barometer) with sampling frequency of 30 and 5 Hz, respectively. The signals obtained were divided into 5-s windows. The dataset used is composed of 25 subjects. ADLs were performed with smartphone worn in 12 different positions. Three sets of features were extracted from each window; spectral domain (maximum/median/fundamental frequencies, max power spectrum, total energy, spectral centroid/spread/skewness/kurtosis/slope/decrease/roll-on/roll-off, curve distance, spectral variation), statistical domain (skewness, kurtosis, histogram, mean, standard deviation, interquartile range, variance, root mean square, median absolute deviation), temporal domain (correlation, temporal centroid, autocorrelation, zero crossing rate, linear regression).

Attal et al. [74] presented various classification methods for HAR. Three accelerometers were placed at the subject chest, right shank, and left ankle. Four supervised classifiers k-NN, SVM, GMM, RF and three unsupervised classifiers k-means, GMM, HMM were compared. Sensors raw data and extracted features are used independently as inputs for each classification method. Nine accelerometer signals were collected from three MTx IMUs. For each signal, the following time and frequency-domain features are calculated: eleven time-domain features (mean, variance, median, interquartile range, skewness, kurtosis, root mean square, zero crossing, peak-to-peak, crest factor, and range) and six frequency-domain features (DC component in FFT spectrum, energy spectrum, entropy spectrum, sum of wavelet coefficients, squared sum of wavelet coefficients, and energy of wavelet coefficients). The k-NN gave the best results between the supervised classification algorithms, while the HMM provided the best results between the unsupervised classification algorithms.

Suto et al. [75] presented various feature extraction methods applied in the HAR domain. The quality of the selected features was assessed using feed-forward artificial neural network, k-Nearest Neighbor, and decision tree. The authors applied a time sliding window of 32 samples with 50% overlap between the consecutive windows. This window length spans 1.6 s time interval as the authors used the WARD database [76] that has a sampling frequency of about 20 Hz. The extracted time-domain features are the mean, variance, mean absolute deviation, root mean square, zero crossing rate, interquartile range, 75th percentile, kurtosis, signal magnitude area (SMA), and the min–max values. The extracted frequency domain features include the spectral energy, spectral entropy, spectral

centroid, and the principal frequency. Other kinds of extracted features are the correlation between axes, autoregressive coefficients, and the tilt angle. The results showed that the proposed machine learning and feature selection approach outperformed other classification methods (up to approximately 100% recognition rates). The authors also showed that a few number of sensors (compared to other works) can be sufficient for good classification performance.

3.3 Autoregression

Vector autoregression is an extension of the ordinary least squares (OLS) regression to the multivariate case. A k-dimensional timeseries (k can be the number of inertial sensors times the number of degrees of freedoms of each sensor) can be linearly regressed in time. In other words, the vector value at time t is linearly dependent on the vector values at previous time steps (up to a certain order) plus a Gaussian noise. Formally, this can be written as follows:

$$\mathbf{y}_t = A_1\mathbf{y}_{t-1} + A_2\mathbf{y}_{t-2} + \cdots + A_p\mathbf{y}_{t-p} + \mathbf{b} + \epsilon_t \tag{4}$$

The $\{A_i\}$ matrices represent the linear transformations, $\mathbf{b}$ is a bias/intercept term, and ϵ_t is a white Gaussian noise. This kind of regressions are called *autoregressive* (AR) models in the terminology of timeseries analysis. A key to the validity of these models is that the given timeseries is *stationary* [30], which roughly means that the statistical properties (such as the moments) of the signal remains fixed over time. Even though this might arguably be not the case for many human actions, features based on autoregressive models have been used with reasonable success. The model's unknowns are these matrices, the bias vector, and the covariance matrix of the noise term. These can be estimated using a variant of the ordinary least squares technique. Once these parameters are estimated, they can be taken as the feature vector or the descriptor of the given timeseries.

Abdu-Aguye et al. [77] investigated the effectiveness of this kind of features for activity recognition using four different classifiers, namely, random forests, canonical correlation forests, ProtoNN, and deep forests. They tested their work over three benchmarks: EJUST-ADL-1 [24], HAPT [42], and Daily and Sports Activities [57]. The dimensionality of the timeseries in each dataset depends on the multitude of the inertial modalities used. The empirical study shows the general efficacy of using the autoregressive coefficients as a feature descriptor of the timeseries, even though the statistical stationarities across all the activities given in the datasets are questionable in several kinds of actions. The results though differ widely based on the dataset used as well as the classifier applied. However, random forests can be considered the best economical choice in terms of the achieved predictive performance and their demand of computational resources.

Hassan et al. [78] presented a HAR system using the smartphone IMU sensors and deep learning. Firstly, the relevant features are extracted from the raw sensor data, e.g., mean, median, autoregressive coefficients, etc. Secondly, the features are processed by a Kernel Principal Component Analysis (KPCA) and a Linear Discriminant Analysis (LDA). Finally, the features are trained using a Deep Belief Network (DBN) for accurate activity recognition. The proposed approach outperformed the traditional recognition methods such as multiclass Support Vector Machine (SVM) and Artificial Neural Network (ANN). The system had been examined for 12 different physical activities where the mean recognition rate was 89.61% and the overall accuracy was 95.85%. The other traditional methods achieved at best a mean recognition rate of 82.02% and an overall accuracy of 94.12%.

Khan et al. [15] presented a HAR system based on a 3-axial accelerometer via augmented feature vector and a hierarchical recognizer. At the last level, the activity state (static, transition, dynamic) is classified using statistical features and ANNs. The upper recognition level utilizes the autoregressive modeling of the acceleration signals. The augmented feature vector consists of autoregressive coefficients, signal magnitude area, and the title angle. The resulting feature vector is further processed by linear-discriminant analysis and ANNs to recognize a particular human activity. The proposed activity-recognition method recognizes 3 states and 15 activities with an average accuracy of 97.9% using a single 3-axial accelerometer placed at the user's chest.

San-Segundo et al. [79] adopted the well-known methods successfully applied in the speech processing domain: Mel Frequency Cepstral Coefficients (MFCCs) and Perceptual Linear Prediction (PLP) coefficients. In addition, Relative Spectra (RASTA) filtering [80] and delta coefficients were also evaluated for IMU processing. The proposed HAR system is based on Hidden Markov Models (HMMs) and recognizes 6 activities: walking, walking upstairs/downstairs, sitting, standing, and lying. A publicly available dataset was used, UCI Human Activity Recognition Using Smartphones [81, 82]. This dataset includes several sessions of physical activity sequences from 30 subjects. The dataset had been divided randomly into 6 subsets to apply a sixfold cross validation procedure. The baseline method for feature extraction performed several calculations in the time and frequency domains. For the time domain, the features included the mean, correlation, signal magnitude area (SMA), and autoregressive coefficients. For the frequency domain, the features included the energy of different frequency bands and frequency

skewness. Other features included the angle between vectors. A total of 561 features were extracted to represent each window (i.e., every 2.56 s). The results presented in this paper enhances significantly the error rates of the baseline method. The adapted MFCC and PLP coefficients enhanced HAR and segmentation accuracies while decreasing the size of the feature vector. RASTA-filtering and delta coefficients reduced the segmentation error rate significantly giving the best results: an activity segmentation error rate lower than 0.5%.

3.4 Raw data

Most of the work done in HAR systems are based on supervised learning over derived features, manually or automatically extracted from the sample timeseries. In the following we survey some of the techniques that are deviated from this direction.

Some research work have treated the streamed signals as direct input without any prior preprocessing and/or feature engineering. Most of that work can be considered as automatic feature engineering (representation learning) through the use of the recent advances in deep neural networks. However, some of that work just directly models the temporal aspects of the incoming signal without explicitly learn any feature representation. In other words, the dynamics of the model itself encode the input signal without explicitly transforming (whether manually or automatically) the signal into some feature space. The most prominent example in that direction is the use of *Hidden Markov Models* (HMMs). An HMM is a latent variable model where there is a sequential relationship amongst the latent variables. It is theoretically founded on the theory of Markov stochastic processes, where, such process satisfies the Markovian property that the current state of the system depends exclusively on the previous state, and hence, independent of earlier history.

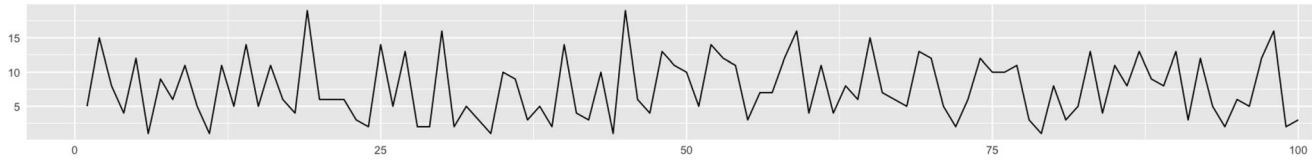
The introduction of latent states, forming the HMM, actually extends this model beyond such Markovian assumption. More specifically, the latent variables represent discrete states that transit among themselves according to some stochastic law. The transition behavior follows the Markovian property, that is, the next transition depends only on the current state. Along side transiting to a new state, a sampled output is observed, through which these hidden states are inferred. The parameters of the model to be learned or estimated are: the transition probabilities, the parameters of the observables distributions, and the distribution of the initial state. By its inherent nature, HMMs are very suitable for modeling sequential data, and in particular, in our context timeseries data. They can be used as inferential engine as well as generative engine. The main advantage is that HMMs are very well mathematically established and hence, can provide explanation to its induced inferences and outputs. They require relatively modest computational resources during training (specially when training using frequentist approaches such as expectation maximization) when compared to deep neural networks.

In the context of timeseries HAR signals, an HMM model can be viewed as encoding the given activity. For example, assume that we have only an HMM H with only two states, that is used to model the timeseries given in Fig. 15a. Figure 15b shows the corresponding states produced by an HMM that models the timeseries and consists of two states. As can be seen from the figure we can perceive the HMM model as encoding the timeseries using a binary string. So, for example, if the timeseries is the z-acceleration of a pushup exercise, then the modeling HMM just encodes this exercise using two levels, they could be the up position and the down position.

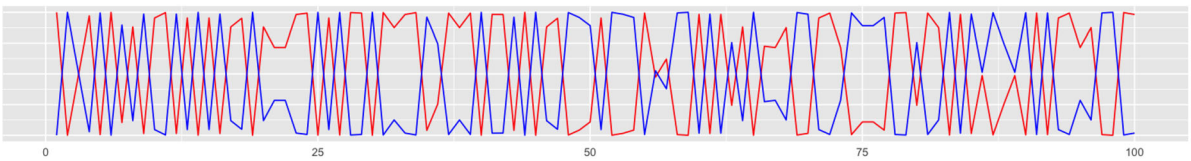
Ashry et al. [83] have used HMM for activity classification. Their main contribution is the novel application of such models, along with devising dissimilarity measures between different HMMs in the domain of activity recognition. Each activity sample, a multi-axial timeseries, is modeled using HMMs for the different axes; *the HMM itself is taken to represent the signal (a descriptor of the signal)*. The training phase consists entirely of deriving the HMM models for all the activity samples in the given dataset. These are considered *templates/prototypes* to be used in the subsequent testing phase. In a 1-Nearest Neighbor fashion, a new test multi-dimensional timeseries is used to induce its characterizing HMM. This HMM is compared against those of the templates using a dissimilarity measure developed by the authors to deduce the closest one whose class of activity is then taken as the decision for the new sample.

Figure 16 shows the architecture of the proposed model in [83]. The authors validated their work on two publicly available datasets, namely, the EJUST-ADL-1 [24] and USC-HAD [12]; and compared the performance against an extant approach utilizing feature extraction and another technique utilizing a deep Long Short-Term Memory (LSTM) classifier. Table 2 shows a snapshot of the results. On the EJUST-ADL-1 they achieved performance metrics (accuracy, sensitivity, specificity, precision, and F-measure) in the order of 90% and on the order of 85% over the USC-HAD dataset.

A crucial notice about this work is that there is no preprocessing step for feature extraction is done, the raw timeseries data streamed from the sensors are fed directly to the HMM. Hence, in some sense, HMM can provide the same advantage as deep neural networks for automatic feature extraction. However, the type of extracted features here are encoded in the form of dynamics. This is very



(a) An sample timeseries, for example, the z-acceleration of a wearable sensor during some action.



(b) An HMM encoding with two states.

Fig. 15 An example signal and its corresponding HMM abstraction with two states

appropriate for low semantical content data such as timeseries. However, this may not be feasible for more complicated data modalities such as visual or audio data. It is also of noticeable interest whether HMMs can be used for transfer learning which is a basic advantage of deep neural networks. In principle it seems plausible that HMMs can be used for that purpose, however, to the best of our knowledge, we have not found any work addressing the feasibility and/or effectivity of HMMs for transfer learning.

On a different track Gomaa et al. [84] apply two well-known statistical and measure-theoretic approaches in the context of human activity recognition. The first approach is time-neglectful in the sense that the temporal order among the sampled points of the timeseries is dropped. It is based on using statistical methods, particularly, *goodness-of-fit* tests, where the timeseries is treated as a collection of unordered points that are generated from some unknown probability distribution. A goodness-of-fit test is a statistical tool used to test the hypothesis that a specific sample (in our case the collection of timeseries points) is generated from a given distribution.

In [84] the authors apply one of the best known goodness-of-fit tests, namely, the *Kolmogorov–Smirnov (KS) test*. This test was developed in the 1930s by two Russian mathematicians Kolmogorov and Smirnov [85, 86]. The second approach is more time-aware considering the ordered nature of the timeseries as a sequential stream of measurements. Their approach is based on direct analysis of the timeseries sample points by using two distance metrics that can measure the dissimilarity between two timeseries even if they have different lengths which is typical in HAR applications. The first dissimilarity metric is based on the estimated autocorrelation coefficients of the given signals. The second is the infamous dynamic-time warping (DTW). The latter is much more accurate, though it is more computationally demanding. Their approach is essentially a 1-NN (1-Nearest Neighbor), where the distance function is taken to be one of these two dissimilarity

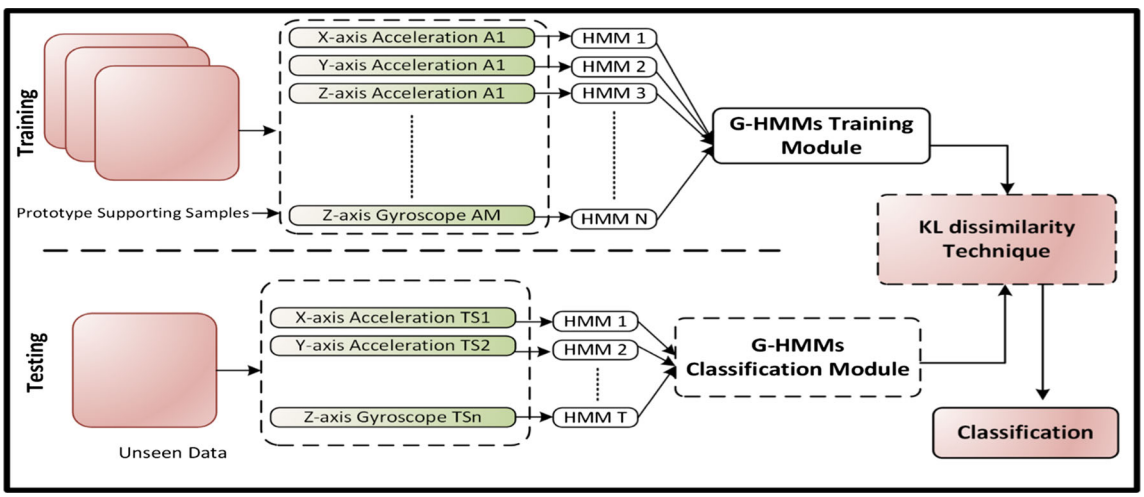


Fig. 16 Framework of the proposed HMM model for HAR [83]

Table 2 A comparison between an HMM-based method [83], LSTM using raw data, and RF [24] over two different public datasets

Dataset	Method (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	*F*-Measure (%)
USC-HAD	RF	78.5	70.45	98.2	60.5	65.1
	LSTM	78.57	72.8	85.71	71	71.88
	HMM	**83.95**	**83.8**	**98.54**	**84.36**	**83.49**
EJUST-ADL-1	RF	81.64	82.47	98.67	84.6	83.53
	LSTM	87.19	78.5	93.1	77	77.74
	HMM	**91.9**	**91.4**	**99.37**	**92.54**	**91.64**

The highest values are highlighted in bold

measures in an infinite dimensional timeseries space. Their experimentation is very limited to one dataset of signals from a wrist-worn accelerometer collected for ADL activity recognition; the dataset is called "Dataset for ADL Recognition with Wrist-Worn Accelerometer" [7, 87], which has 14 activities of daily living. The DTW-based method is more stable reaching an accuracy of about 84% which is better of by more than 25% that of the other autocorrelation dissimilarity measure. The results based on the KS-test are much less promising.

On a similar track, where the sequential nature of the timeseries signal is ignored, Gomaa et al. [88] treat a timeseries as an unordered set of measurements sampled from an unknown probability distribution. They choose some random timeseries samples from each activity and fit them using probability distributions. These templates are then used as prototypes against which a new inertial motion sample is tested to determine the corresponding action. The authors used two probabilistic modeling techniques: *histograms* and *kernel density estimates*. For histograms, Manhattan distance is used as a measure of dissimilarity, whereas Manhattan distance and KL-divergence are used as measures of dissimilarity for kernel density estimates. Again, the validation of the approach is very limited by experimenting only on one dataset, which is the "Dataset for ADL Recognition with Wrist-Worn Accelerometer" [7, 87]. This dataset has 14 ADLs. For some configurations they managed to reach above 80% accuracy.

3.5 Embedding

All of the previous discussion have focused on either handcrafted feature engineering or using the timeseries data directly. A third alternative has gained much attention recently with the emergence of deep learning, in particular, deep neural networks, which is *representation learning*, or more concretely *automatic feature extraction*. Generally, automatic feature extraction through deep neural computing has achieved tremendous success in recent years and allowed for technological applications, and concerns as well, that had been unimaginable before. As the burden of engineering effective features is now alleviated, it came with a price: we don't understand the generated embedding/representation, and generally the inferred decisions. In some cases, specially with the visual modality, some work has been done to visualize and interpret the resulting hierarchical representations. Generally, the ability to generate informative embedding is crucial for effective transfer learning. In addition, the data manifold in the embedding space is generally easier to manipulate than the original input space.

Deep neural networks entered the mainstream with the seminal work of Hinton and Salakhutdinov [89]. Since then, the general trend in machine learning research has been directed towards the adoption and analysis of deep methods as applied in different domains. Abdu-Aguye et al. [90] train a convolutional neural network (CNN) essentially for representation learning to generate embeddings for inertial timeseries data that are then used for transfer learning across different datasets. They train a CNN and use its convolutional filters as a *feature extractor*, then subsequently they train a feedforward neural network as a classifier (or for that matter any classifier can be used instead) over the extracted features/embeddings for other datasets. The overall architecture of their system is shown in Fig. 17.

The authors use a *spatial pyramid pooling* layer in order to generate a fixed-size embedding regardless of the length of the input timeseries, hence, alleviating the need to fix the input length of the timeseries. They validate the effectiveness of their embedding model on five different activity recognition datasets: EJUST-ADL-1 (aka Gomaa-1) [24], HAPT [42, 91], EJUST-ADL-2 (aka HAD-AW) [14], Daily and Sports Activities [57], and REALDISP [13].

Tables 3 and 4 show example results from the transfer learning, the first table shows the performance metric in terms of the overall accuracy, and the second shows the computational resources in physical time (seconds). The diagonal is the baseline (self-testing), whereas cross-testing are the off-diagonal entries.

The machine used to obtain these timings utilized an 8-core Intel Xeon Platinum 8168 CPU with 16 GB of RAM. During the training cycles, CPU-only training was used to give a fair estimate of performance and eliminate any extraneous advantages due to the use of graphics processing units (GPUs). The authors thereafter report the mean and standard deviation of the obtained times measured in seconds.

In the latter, a model is pre-trained on each source dataset mentioned in the first column, and then fine-tuned and tested on a target dataset in each corresponding row in the table. It is apparent that transfer learning, and hence, the generated embeddings, are effective where the maximum reduction in accuracy is about 6%, and in most cases it is about 3%. On the other hand, the bottom table shows tremendous saving in computational resources as a result of transfer learning. The only computation here is required for fine-tuning. In the worst case (i.e., on the Daily Sports dataset) about 24× speedup is achieved compared to the self-testing scenario; and about 52× speedup in the best case (i.e., on the REALDISP dataset). It is also apparent that when the authors used the datasets, Daily Sports and EJUST-ADL-2, the best results have been achieved regarding transfer learning. This means that their pre-trained model is the most powerful when transferring the gained knowledge from these datasets to another dataset. This can be attributed to the wide spectrum of activities contained in these two datasets (19 actions in the former and 31 in the latter); see Table 15 for full description of these datasets.

Another work based on proper embedding representation learning of timeseries inertial signals is done by Abdu-Aguye et al. [92]. In particular, they employed such embedding representation in a deep *metric learning* (aka *similarity learning*) framework for activity recognition. A *deep Triplet Network* is used to generate fixed-length embeddings/descriptors from activity samples. Such embeddings are located in a classical Euclidean space and are to be used for activity classification. The overall system architecture is similar to that shown in Fig. 17, except for the use of triplet network with the corresponding loss function and the replacement of the fully connected layers towards the end of the information flow by a 1-Nearest Neighbor classifier with the Euclidean distance as the dissimilarity metric. The triplet loss can be defined as follows:

$$\mathcal{L}(\theta) = \sum \max\left(\|E_{\mathrm{a}} - E_{\mathrm{p}}\|_2^2 - \|E_{\mathrm{a}} - E_{\mathrm{n}}\|_2^2 + \alpha, 0\right) \quad (5)$$

where E_{a} is the embedding of the *anchor* sample, E_{p} is the embedding of the *positive* sample, and finally, E_{n} is the embedding of the *negative* sample. The summation of the loss function is taken over all the triplets in the training set. α represents a margin parameter that describes the minimum desired spacing between entities of different classes.

The authors evaluate their work using the same 5 datasets used in [90]. Transfer learning (cross-testing) is also performed across different datasets: a model is built by pretraining in some source dataset, then fine-tuned and tested on a target dataset. Table 5 shows a sample of the results, both self-testing and cross-testing, of the proposed deep metric architecture of the five mentioned datasets. Comparing these results with those of the previously mentioned work, Table 3, we observe the following; (1) the results of self-testing are comparable, both methods are in close proximity to each other, however, (2) the method proposed in [90] is much more capable to transfer to new datasets, that is, the induced embeddings from this scheme is much more abstract representation of the time-series inertial signals that can transcend the source dataset. As the two lines of work do their evaluation over the same list of datasets, it is worth to investigate what factors affect the transferability power of the underlying model.

The work done in [90] learns the embeddings within an end-to-end training pipeline for the overall final purpose of activity recognition, whereas the work in [92] learns the

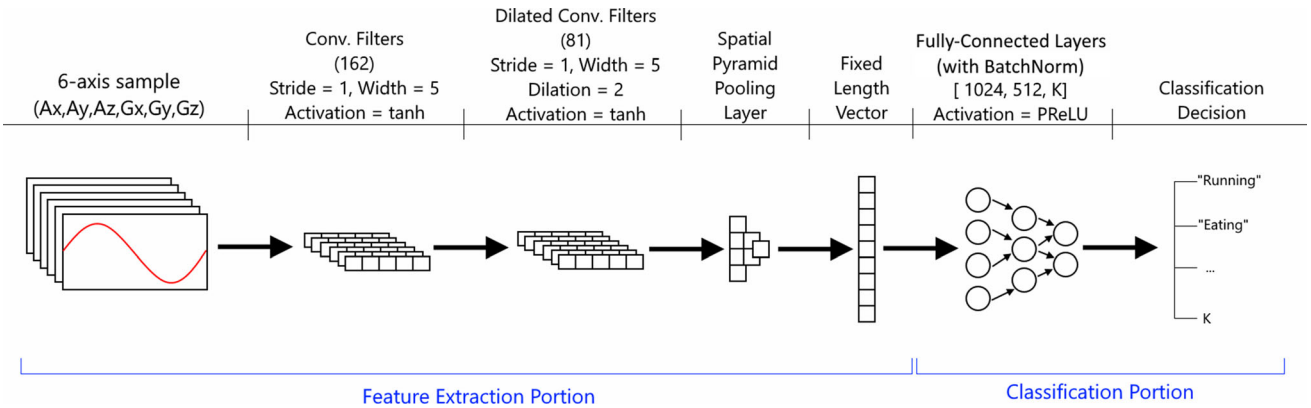


Fig. 17 CNN for feature extraction + classification [90]

Table 3 Classification accuracy using 4-2-1 pooling [90]

Training dataset	Testing dataset				
	Gomaa-1	HAPT	D_Sports	HAD-AW	REALDISP
Gomaa-1	**96.99 ± 1.28**	94.53 ± 1.55	92.19 ± 0.64	82.66 ± 1.01	71.94 ± 2.64
HAPT	95.09 ± 1.90	**96.00 ± 1.18**	91.06 ± 0.76	83.10 ± 1.12	68.85 ± 1.77
D_Sports	95.98 ± 1.48	95.24 ± 1.09	**95.31 ± 0.36**	83.20 ± 1.15	71.14 ± 3.70
HAD-AW	95.93 ± 1.55	95.61 ± 1.18	92.38 ± 0.50	**88.36 ± 1.18**	71.63 ± 2.15
REALDISP	95.58 ± 1.18	94.82 ± 1.55	92.59 ± 0.41	83.60 ± 1.02	**76.60 ± 1.88**

The highest values are highlighted in bold

Table 4 Training times (in seconds) for self-testing and cross-testing scenarios (4-2-1 pooling) [90]

Training dataset	Testing dataset				
	Gomaa-1	HAPT	D_Sports	HAD-AW	REALDISP
Gomaa-1	**426.10 ± 5.95**	16.44 ± 0.50	62.30 ± 1.37	49.70 ± 1.18	27.07 ± 0.95
HAPT	9.91 ± 0.29	**513.33 ± 4.50**	62.02 ± 0.94	49.44 ± 0.90	27.02 ± 0.65
D_Sports	9.98 ± 0.40	16.30 ± 0.40	**1503.26 ± 13.38**	50.20 ± 1.35	27.08 ± 0.75
HAD-AW	9.92 ± 0.35	16.32 ± 0.54	63.12 ± 1.63	**1952.15 ± 16.19**	27.16 ± 0.63
REALDISP	9.98 ± 0.31	16.33 ± 0.45	62.57 ± 1.42	49.84 ± 1.26	**1413.45 ± 16.13**

The highest values are highlighted in bold

embeddings in a more abstract setting, where the task of activity recognition appears more implicit in the choice of the three examples for the triplet loss: the anchor, the positive example, and the negative example. It may as well be the technique used for action classification may play a role in effecting the transfer learning; in the former work a fully connected neural network is used, whereas in the latter a 1-Nearest Neighbor is used. Maybe the choice of 1 neighbor leads to overfitting and more smoothing are needed as a regularization. All of these need to be investigated, as transfer learning in general, and its particular use in inertial timeseries data are so crucial in the success of such systems and the practical widespread effective use of wearable devices.

Khaertdinov et al. [93] use a variant of deep metric learning equipped with attention for HAR. The main focus of the paper is to design a system that is robust against two problems in wearables-based HAR systems: (1) inter-class similarities (different activities have similar inertial motion patterns) and (2) subject heterogeneity, where different subjects perform the same activity quite differently. The authors show the detrimental effect of these two factors on the recognition performance by using t-SNE visualization over 2-dimensional Euclidean space. The crux of the argument of the plotted graphs is that different actions have quite overlapping clusters and the same action may spread over different clusters in that 2-dimensional space. These motivate the whole work, however, it can be criticized that visualization is a quite poor evidence as it projects from the rich high dimensional space of the timeseries into very low 2-dimensional space, with much less degrees of freedom.

More evidence is needed, for example, plotting 3-dimensional visualizations, plotting multiple 2-dimensional visualizations over varying features, or any other means. The main technical contributions of the paper behind that motivation of well-class separability are: (1) developing a hierarchical triplet loss function and (2) developing a smart triplet mining algorithm. The *hierarchical triplet loss* is based on a data structure called the "hierarchical tree" that stores information about the inter-class and intra-class

Table 5 Classification accuracy of embeddings using deep metric learning [92]

Training dataset	Testing dataset				
	Gomaa-1	HAPT	D_Sports	HAD-AW	REALDISP
Gomaa-1	**96.69 ± 1.23**	89.60 ± 1.72	91.34 ± 0.41	68.67 ± 1.09	59.48 ± 2.77
HAPT	83.27 ± 2.16	**96.25 ± 1.09**	89.98 ± 0.74	64.46 ± 1.46	50.61 ± 2.19
D_Sports	86.16 ± 2.69	88.67 ± 1.17	**96.59 ± 0.39**	68.62 ± 1.14	56.59 ± 2.27
HAD-AW	90.75 ± 1.42	89.10 ± 1.41	91.35 ± 0.56	**85.28 ± 1.19**	60.60 ± 2.00
REALDISP	91.16 ± 2.14	91.40 ± 1.38	91.63 ± 0.47	70.10 ± 1.65	**75.01 ± 1.92**

The highest values are highlighted in bold

distances of all activity classes in the dataset. Given two activity classes a_1 and a_2, their inter-class distance can be computed as follows:

$$d_{\text{inter}}(a_1, a_2) = \frac{1}{\|a_1\| \cdot \|a_2\|} \sum_{i \in a_1, j \in a_2} \|E_i - E_j\|_2^2 \tag{6}$$

where E_k is the embedding of sample k. So, it is the average of all distances between pairwise samples from the two classes. On the other hand, the intra-class distance for activity b can be computed as follows:

$$d_{\text{intra}}(b) = \frac{1}{\|b\|(\|b\| - 1)} \sum_{i,j \in b} \|E_i - E_j\|_2 \tag{7}$$

Each leaf node corresponds to a certain class label. Classes are merged at different levels of the tree based on some merging threshold. Now the hierarchical loss function can be computed as follows:

$$\mathcal{L}(\theta) = \frac{1}{2N} \sum_{i=1}^{N} \max\left\{ \|E_i^{\text{a}} - E_i^{\text{p}}\|_2^2 - \|E_i^{\text{a}} - E_i^{\text{n}}\|_2^2 + \alpha', 0 \right\} \tag{8}$$

where N is the number of triplets, $E_i^{\text{a}}, E_i^{\text{p}}, E_i^{\text{n}}$ are the embeddings of the anchor, positive, and negative samples in the triplet i. α' is the dynamic margin that depends on the intra-class distance of the anchor activity as well as the level ℓ in the hierarchical tree at which the anchor and negative classes are merged. So, adapting the margin to the particular activities involved in the triplet makes the training more robust and resilient to the two opposing effects of inter-class similarities and subject heterogeneity. The second main technical contribution of this work is the technique for mining triplets for a mini-batch. They call their method "anchor-neighbor sampling".

It is essentially a smarter guided selection of the triplets to make proper use of the constructed hierarchical tree to improve the training process and well separate the classes. The authors evaluate their work over three public benchmark datasets: PAMAP2 [94], USC-HAD [12], and MHEALTH [95]. Their embedding model is based on LSTM with two types of attention, attention based on the temporal sequence of timeseries points and spatial attention scanning multiple channels (different inertial types and axes) at the same time. They report results over all three datasets and multiple configurations of their model (for example, use one or both types of attention) and compare with other state-of-the-art methods.

Strömbäck et al. [31] create an iconic dataset, called "MM-Fit" for workout exercises. This dataset is multimodal collected from varying devices streaming the following types of data: (1) inertial motion data collected from multiple devices including smartphones, smartwatches, and earbuds, (2) multi-view RGB-Depth video streaming, and (3) 2D and 3D pose estimates. The task is to do activity recognition, particularly in this case, recognition of the workout exercise. The authors compare the power of different deep recognition models based on unimodal vs. multimodal data types.

Each modality is trained separately using a stacked convolution autoencoder to produce an embedding representation of the crucial features from the given modality. For the inertial motion timeseries the autoencoder architecture consists of 1D convolution filters. As for the visual streaming, they use the 3D pose estimates from [96] and the relative joint positions are encoded using cylindrical coordinates as such coordinate system is more robust against background and illumination variation. Such 3D poses are arranged into 3D images that are fed to an autoencoder that is based on 2D convolution filters.

Each unimodal type of sensory data is trained independently to generate an embedding that is output from the later layer of the autoencoder (the last layer of the encoder part). These induced embeddings representations of the different modalities are then fused and fed to a multi-modal fully connected autoencoder. Lastly, the fused embeddings of this latter network are used to train a fully connected classification network to recognize 10 different workout exercises. Their multi-model framework achieves 96% accuracy on unseen subjects in their dataset MM-Fit. This is in contrast with 94% using data from the smartwatch only, 85% from the smartphone only, and 82% on data from the earbud device. A general critic to this work is that their proposed model architecture maybe overly complicated.

Mahmud et al. [97] proposed a self-attention model that utilizes different types of attention mechanisms to generate higher dimensional feature representation. The authors evaluate on four publicly available HAR datasets: PAMAP2 [94], Opportunity [98], Skoda [99], and USC-HAD [12]. The sensor attention maps can capture the importance of sensors modalities and location in predicting the activity classes. The authors adopted the self-attention architecture from the Neural Machine Translation (NMT) task [100] into the HAR problem and proposed a model incorporating self-attention with sensor and temporal attention.

The authors proposed that the data samples are equivalent to words and the time windows are analogous to sentences. They introduced the first attention layer on the raw input. Secondly, the authors adopted the self-attention and positional encoding from the transformer architecture into the HAR problem in order to capture the spatio-temporal dependencies of sensor signals and their modalities. After a number of self-attention blocks, the authors added another layer of attention to learn the global attention from

the context. Finally, a fully connected layer was placed to classify the activity. The proposed model achieved remarkable performance enhancement over the existing state-of-the-art models on both benchmark test users and Leave-one-user-out evaluation. The authors also observed that the sensor attention maps was able to focus on the impact of the sensors modality and placement in predicting the various activity classes.

Tao et al. [101] proposed an attention-based model for HAR using multiple IMUs worn at various body locations. A feature extraction module extracts the most discriminating features from every sensor with CNNs. An attention-based fusion method learns the sensors importance at various body locations and generates an attention-based feature representation. A feature extraction module learns the inter-sensor correlations that are connected to a classifier predicting classes of activities. The proposed method was evaluated using 5 public datasets: Daily [102], Skoda [103], PAMAP2 [94], Sensors [67], DaphNet [104]) and outperformed the state-of-the-art approaches on various activity categories.

Li and Wang [105] proposed a deep learning method that is based on residual blocks and bi-directional LSTM (BiLSTM). The model extracts spatial features of multi-dimensional signals from IMUs using the residual block. The model then gets the forward and backward dependencies of the features sequence using BiLSTM. The resulting features were input to a Softmax layer to perform the HAR task. The optimal parameters of the model were acquired experimentally. A new dataset was proposed containing six activities: sitting, standing, walking, running, going upstairs, going downstairs. The proposed model was evaluated on the newly proposed dataset and two other public datasets namely WISDM [106] and PAMAP2 [94]. The accuracies achieved are 96.95%, 97.32% and 97.15% on the newly proposed dataset, WISDM and PAMAP2, respectively.

In [107], the authors proposed a graph neural network (GNN) that is end-to-end. This system captures the sample information with an efficient way and also captures the relation with the other samples as an undirected graph form. This is probably the first contribution where time series are represented as a graph-based structural representation aiming at HAR utilizing sensor data. The proposed approach was experimented on six publicly available datasets. The proposed method achieves for all the datasets approximately 100% recognition accuracy. The GNN classification was done at two steps. First, the node classification was done for each node $v \in V$ of G that is labeled as y_v to predict the labels of every node v by calculating an embedding vector h_v where $y_v = f(h_v)$ and f is a differentiable function for approximation. f was utilized in order to make mapping for the similarity between the nodes in an embedding dimension with the real graph nodes that are existing in the real graph. In the embedding dimension, some information was saved in a compressed approach withing h_v. Second, the graph structure is considered as well for the graph classification beside using the node information.

3.6 Input tensor structure

In the analysis of inertial motion data, a diverse set of tensor structures have been used as inputs to the learning architectures (whether classical or variants of deep neural networks). The input tensor structure essentially depends on several factors, notably, the number of IMU sensors to be used for analysis, the number of axes of the sensor(s) used, whether vectorization (concatenating into a single vector) is used, whether raw sensor data or some sort of extracted features are used, etc.

For example, one may use only the accelerometer data in one dimension (or take the magnitude of several directions into one resultant vector). In this case 1D tensors are used. It may be the case of taking the three axes of an accelerometer and concatenate them together into a single vector (vectorization), making also for a 1D input tensor. One may take the three axes of the accelerometer making each in a separate channel, making 2D tensor input, or making them in one channel, however, stacked vertically, making again a 2D tensor. One may use several sensors: accelerometer, gyroscope angular velocity, and magnetometer. Each has three axes that can be stacked vertically in a single channel. Thus, having three channels, each is a 2D array of three axial readings, so in this case, we have 3D input tensor. So, in conclusion the input tensor structure depends on the particular work and the authors' choice.

An example of such variations can be found in the work of Mubarak et. al. [90], where the authors used six 1D channels: accelerometer-x, accelerometer-y, accelerometer-z, gyroscope-x, gyroscope-y, and gyroscope-z. Each of these is a 1D timeseries, so the whole tensor can be considered a 2D tensor input. Similar strategy was applied in the work of Abeer et. al. [32]. In [83] the authors use hidden Markov models (HMMs) to model inertial motion data. For each axis of each sensor a dedicated HMM is used to model the dynamics of motion along this axis. So, the input to each HMM is a univariate timeseries motion data, in other words, it is a 1D tensor. On the other hand, Madcor et. al. [18] have used two sensors, the accelerometer and gyroscope, each sampled in three axes. The timeseries of the six axes are, after some preprocessing, concatenated into one big vector (vectorization) to be input to the learning architecture. So, here the input is a 1D tensor.

3.7 Transformers

In this subsection, we survey the use of the transformer model in Human Activity Recognition from IMUs time series data. The transformer model is a deep learning NN model that was developed primarily for the NLP and computer vision tasks.

In [108], the authors proposed a one-patch lightweight transformer that combines RNNs and CNNs advantages. In addition, the authors proposed TransFed that considers more the privacy concerns. TransFed is a federated learning classifier that utilized the proposed lightweight transformer. The authors designed a test framework to acquire a new HAR dataset gathered from 5 subjects, and then utilized the novel dataset in order to assess the performance of the HAR classifier in the two settings of federated and centralized environments. In addition, the authors used a public dataset to assess the performance of the HAR classifier in centralized environment for comparison with other HAR classifiers. The results showed that the proposed approach outperformed the state-of-the-art HAR classifiers which are based on CNNs and RNNs and in the same time was more efficient computational-wise. Table 6 presents the hyper-parameters adopted during the training phase.

Table 7 presents a comparison of 2 main features between the proposed transformers in [108] with the RNNs and CNNs. The RNN models have no parallelization during training due to the sequential nature, thus the model is slow and computationally expensive. The CNN models can have parallel computation, however, computationally extensive due to the convolution function. The proposed transformer approach removes the recurrence and convolution. That are replaced with a self-attention approach to perform input/output dependencies. The method relies completely on an attention mechanism to calculate input/output representations. Moreover, the transformers allow for computations parallelization. Also, RNNs and CNNs utilize a huge number of parameters (hundreds of thousands or more), while the proposed transformer in [108] has only 14,697 parameters. In addition, the proposed transformer utilizes a single patch not multiple-patches, thus, it is computationally efficient.

Table 8 presents the comparison of the proposed transformer in [108] with state-of-the-art approaches. The proposed approaches achieved a big enhancement in accuracy: (1) trained and tested on the proposed new dataset (labelled with b) and (2) the WISDM dataset (labelled with a). The proposed approach has an accuracy of 98.74% (federated setting) and 99.14% (centralized setting) on the newly gathered dataset. The proposed approach has an overall accuracy of 98.89% on the WISDM dataset (centralized setting).

Table 6 Hyper-parameters of each subject for federated learning [108]

Hyper-parameter	Value
Learning rate	0.01
Number of epochs	100
Batch size	30
Weight decay	0.001
Transformer layers	2
Multi-attention heads	5
Input shape	140×9

Table 7 Comparison of the transformer method with the RNNs and CNNs methods based on the computational costs [108]

Method	Parallelization	Computationally expensive
RNNs	No	Yes
CNNs	Yes	Yes
Transformers	Yes	No

In [112], the authors proposed Human Activity Recognition Transformer (HART) a transformer framework based on sensors adapted to the IMUs inside the mobile devices. The results on HAR activities on many public datasets revealed that HART has less Floating point Operations Per Second (FLOPS) and fewer parameters achieving better results than the state-of-the-art methods. In addition, the authors provided performance assessments on various architectures in heterogeneous environments. The proposed models can generalize better either on various sensor devices or on-body locations.

Table 9 shows the F-score for various models on four datasets: RealWorld, HHAR, MotionSense, and SHL. Moreover, the authors combine five datasets to form the largest one with biggest diversity for assessment. The combination had thirteen different activities predicting data with high-class imbalance from various devices and locations.

For the proposed configuration with MobileViT and MobileHART, 2 network settings were developed: extra small (XS) and extra extra small (XXS). The XS and XXS are differently characterized by the filters number, the expansion factors leading to filter growth after every MV2 layer, and the MobileViT / MobileHART blocks embedding size. The layers number and attention heads are the same in the 2 network settings.

Table 8 Comparison of the transformer method with state-of-the-art approaches for HAR classification [108]

Scheme	Centralized or federated	Number of activities	Accuracy (%)
[109][a]	Federated	6	89.00
[110][a]	Centralized	6	97.63
[111][a]	Centralized	6	96.70
Proposed[a]	Centralized	6	98.89
Proposed[b]	Federated	15	98.74
Proposed[b]	Centralized	15	99.14

In [116], the authors proposed a self-attention Transformer model for ADLs classification. The authors compared the proposed method with a recurrent Long-Short Term Memory (LSTM) model. The proposed approach is a two-level hierarchical model where atomic activities are detected in the first step then in the second step the probability scores are calculated and used by the Transformer-based to classify 7 more complex ADLs. The experiments showed that the Transformer model was on bar with and sometimes outperforms LSTM in the subject-dependent setting (73.36% and 69.09%) only depending on the attention-approach to present global dependencies between input and output without using any recurrence. The proposed model was evaluated on 2 different segment lengths proving its effectiveness to learn long-range dependencies of actions of short lengths occurring in complex activities.

In [117], the authors presented a Two-stream Convolution Augmented Human Activity Transformer (THAT) model. The proposed approach utilized a two-stream architecture to extract both time-over-channel and channel-over-time features, then use the augmented multi-scale convolution transformer to extract the range-based patterns. Results on four evaluation datasets showed that the proposed approach is more effective and efficient compared to the state-of-the-art models. Table 10 presents the THAT model recognition accuracy in comparison with five state-of-the-arts models on four evaluation datasets.

In [121] the transformer model was utilized for Human Activity Recognition from time-series signals. The self-attention approach in the transformer model present the individual dependencies between time-series values. The attention mechanism has a comparable performance to the state-of-the-art CNN-LSTM (e.g., [122]). The transformer model was evaluated on the KU-HAR dataset spanning a wide range of activities [123] achieving an average recognition accuracy of 99.2% compared to 89.67% of the Random Forest with FFT [123].

4 Confined versus streamed actions

Most of the work done in HAR target offline applications and hence, focusing on recognizing confined actions, that is, the timeseries sample is assumed to represent exclusively one action/activity of interest. A line of research though has been targeting online real-time applications, and hence, focusing on recognizing a continuous stream of actions in real-time. Most of this work, though, use vision-based models [124], in which actions in a video stream are detected and recognized in both offline and online modes [125]. In other approaches, data are collected from distributed ambient sensors such as microphones, cameras, and motion sensors that are attached to fixed locations in the surrounding environment (e.g., walls, cupboards doors, microwave ovens, and water taps). Examples of these system include the Aruba and Tulum datasets [126]. These are created within the CASAS smart-home project [127]. Studies on these datasets are done by Yala et al. [128].

Table 9 The F-scores on five datasets [112]

Architecture	UCI	MotionSense	HHAR	RealWorld	SHL	Combined
CNN [110]	94.53	96.91	96.99	91.71	87.08	86.87
CNN-LSTM [113]	92.79	97.80	96.97	92.33	88.92	88.03
ViT [114]	93.66	98.09	97.39	94.57	93.11	93.52
HART [112]	94.49	98.13	98.31	95.49	94.39	**94.59**
$\text{HART}_{\text{OneMSA}}$ [112]	94.37	98.29	98.28	95.74	**94.88**	94.07
MobileViT (XS) [115]	96.89	98.19	98.65	95.07	94.14	93.19
MobileHART (XS) [112]	97.20	**98.49**	**98.72**	**95.81**	94.86	94.24
MobileViT (XXS) [115]	96.55	98.42	98.30	93.72	91.46	91.07
MobileHART (XXS) [112]	**97.67**	98.32	98.19	94.32	92.60	91.91

The highest values are highlighted in bold

Table 10 The THAT model recognition accuracy in comparison with five state-of-the-arts models on four evaluation datasets [117]

Methods	Office	Activity	Meeting	Activity + meeting
S-RF [118]	75.3	80	84.7	82.6
S-HMM [119]	79.7	75	83.4	80.5
LSTM [118]	91.4	89.7	90.6	90.1
CNN [118]	96.4	94.3	96.2	95.4
ABLSTM [120]	97.1	95.6	96.8	95.9
THAT [117]	98.2	98.4	99.0	98.6

These sensors, however, have many restrictions owing, in particular, to their fixed nature, and user activities are not detectable, let alone recognizable, if the user goes outside the spatial area where these sensors are installed. Additionally, from the perspective of privacy, it is neither acceptable nor convenient to monitor people continuously using cameras in private rooms. And even for less private places, monitoring is susceptible to security breaks.

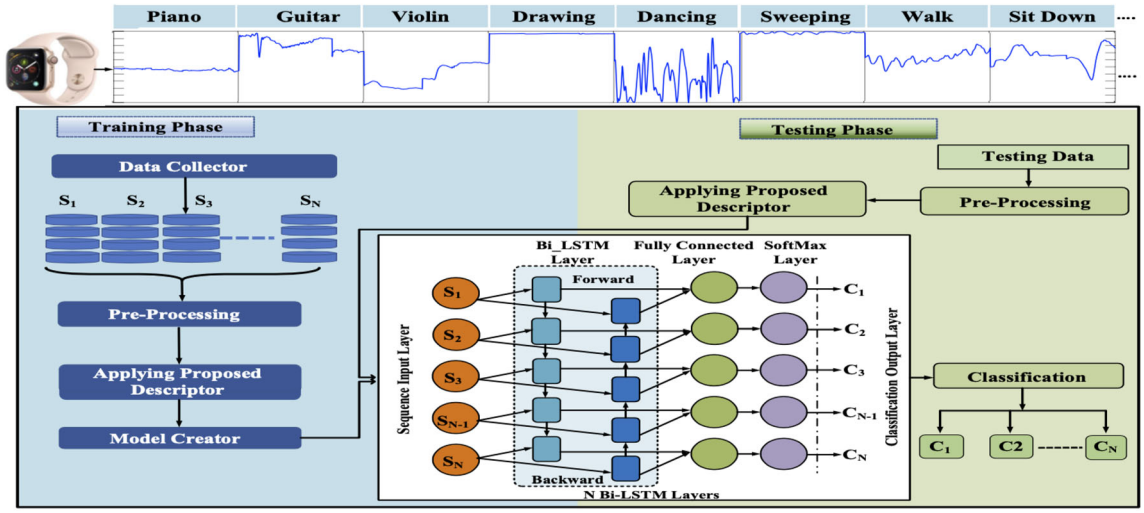
For detecting activities in a continuous stream of inertial motion signals from wearable devices, some segmentation approaches such as activity-based windowing, time-based windowing, and sensor-based windowing [128] have been proposed. However, each one of them admits some kind of drawback. In activity-based windowing, the activities are in general not well distinguished resulting in largely imprecise activity boundaries. In time-based widowing, streaming data are divided into fixed-time intervals/windows. However, it is faced with the problem of determining the proper window length. If it is too small, the window may contain insufficient information for decision making; if it is too large information of multiple activities may be included in the allotted window. Generally, this parameter is problematic as it may differ according to the activity, the gender of the subject performing the activity, the age of the subject performing the activity, the sampling rate of the IMU device, etc. So, smarter techniques should be used to determine an adaptive window size, for example, detecting the change in the stochastic properties of the signals, as generally is done in Markov regime switching-based methods. In sensor-based windowing, each window contains the same number of sensor events [128]. A major drawback of this method is that the window may contain sensor events that are widely separated in time, and it may contain sensor events of more than one subject as in the Tulum dataset.

One recent work that addresses the continuous recognition of streaming of actions is done by Ashry et al. [129]. They process inertial motion signals streamed from IMU sensors on-board a smartwatch. They collect their own dataset, called "CHAR-SW", which consists of different streams of activities of daily living, see Fig. 18a. See Table 15 for a full description of the dataset. They used cascading bidirectional long short-term memory (Bi-LSTM) for their analysis. Bi-LSTM has better and more stable performance than the uni-directional LSTM as the former can perform smoothing as at any point in time the past and future motion samples are available for the current prediction. From another perspective this architecture may not be feasible for real-time systems where only past samples (and possibly a small window of future samples) can only be available. In this work the input to the Bi-LSTM is not the raw timeseries data, however, it is a descriptor which composed of a continuous stream of the following extracted windowed features: autocorrelation, median, entropy, and instantaneous frequency. So their approach is a mixture of classical feature engineering, as manifested by the handcrafted descriptor of the input timeseries, and deep automatic feature extraction, as manifested by the Bi-LSTM which automatically captures the sequential dependency of the streaming actions. So, the technique can be viewed as converting the original timeseries into a coarser more abstract timeseries. The system architecture is shown in Fig. 18b. The evaluation of the system showed that it can recognize activities in (almost) real-time with an accuracy up to 91%. Figure 19 shows a framework for gait-based person identification [20].

Khannouz and Glatard [130] proposed data stream classifiers on the HAR use case. The authors measured the classification accuracy and resource consumption (in terms of runtime, memory, and power) of five stream classifiers on two real human activity datasets [13, 38] and three synthetic datasets. The results showed that the Hoeffding Tree, the Mondrian forest, and the Naive Bayes classifiers were overall better than the Feedforward Neural Network and the Micro Cluster Nearest Neighbor classifiers on four datasets including the real ones. The three best classifiers overall performed much worse than an offline classifier on the real datasets. The Hoeffding Tree and the Mondrian forest are the most memory draining with longest runtime but no difference in the power consumption between classifiers was reported. The stream learning for HAR on connected devices had two main challenges: high memory consumption and low $F1$ scores.

Krishnan and Cook [131] proposed a sliding window technique for HAR in an online streaming behavior which means classifying activities whenever new sensor events came in. Since different activities can be best classified by different window lengths of sensor events, the authors incorporated the time decay and mutual information-based weighting of sensor events within a sensor window. More contextual information like the previous activity and the previous window activity were also added to the feature characterizing a sensor window. The experiments

(a) Snapshot streaming actions of CHAR-SW dataset



(b) CHARM-Deep system architecture

Fig. 18 Online analysis of a streaming sequence of actions using IMU sensory data from a wearable device [129]

evaluating these methods on real-world datasets concluded that combining mutual information weighting of sensor events and the past contextual information into the feature leads to the best performing streaming activity recognition.

Abdallah et al. [132] proposed a dynamic human activity recognition framework with growing data streams. This framework was called STAR standing for STream learning for mobile Activity Recognition. This framework comprised of incremental and active learning methods for real-time recognition and adaptation in streaming situations. While the data stream grows, the authors refine, improve, and personalize the model to handle the drift in the given data stream. To detect the concurrent activities, the authors applied an online clustering for each data chunk. The prediction and adaptation methods were deployed on each cluster. The experimental results using three real activity datasets, OPPORTUNITY [98], WISDM [106], Smart Phone Accelerometer Data (SPAD) [133], showed that the proposed approach enhanced the performance of the activity recognition specifically across different subjects.

Abdallah et al. [134] proposed and evaluated a method for concept evolution (e.g., new action or activity) that was applied to emerging data streams. This method continuously screens the streaming data movement in order to detect any evolving updates. This method was able to detect the arrival of any new concepts either they were normal or abnormal. This approach has also applied continuous and active learning to adapt with the observed concepts in real-time. The authors evaluated the proposed approach on the HAR datasets OPPORTUNITY [135] and WISDM [106] as benchmarks for the emerging data streaming application. The proposed approach proved its

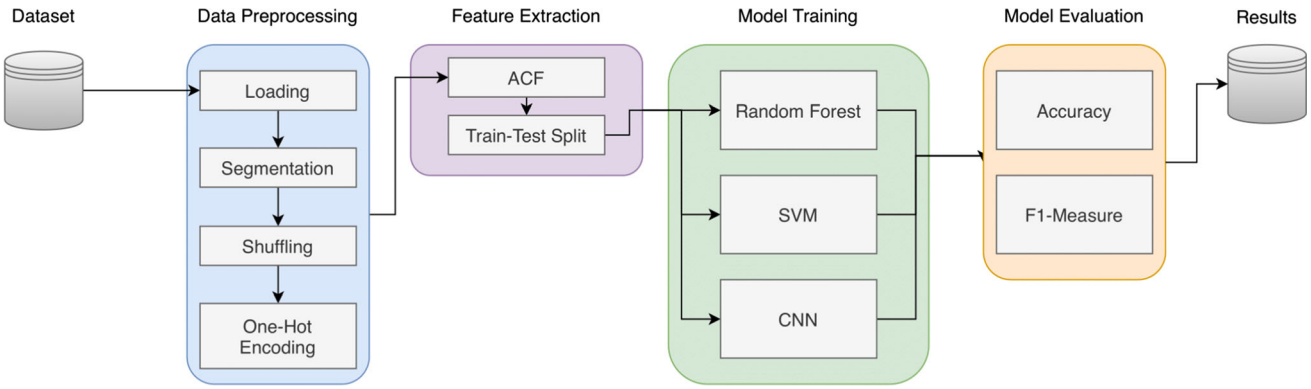


Fig. 19 Framework for gait-based person identification [20]

effectiveness on these datasets in observing new concepts and adapted continuously with low computational cost.

Yala et al. [136] proposed an approach of HAR using online sensor data. The authors presented four techniques for extracting features from sensor events sequence. They proposed a sensor window approach to perform HAR online by recognizing activities when new sensor event occurs. As different activities can be better identified by various window lengths, the authors proposed a mutual information-based weighting of sensor events within a window. As some activities do not necessitate many movements, the authors proposed a "last-state" of sensor feature set within a window to identify activities. These techniques are evaluated on first 6 weeks of Aruba dataset and on 3 months of Tulum dataset [126]. These methods showed an enhancement in classification accuracy over the baseline approach when removing events with missing labels. In the baseline case, the authors construct a feature vector with fixed dimension that contains the first and last sensor events time, the window duration, and the different sensor events count within the window.

Zhang and Ramachandran [137] adopted an online method called Very Fast Decision Tree (VFDT) to simulate the real HAR scenario. The two main contributions are (1) the authors trained the model online using the data samples only once for training and (2) the model can be updated to recognize new activities after building the VFDT by adding small number of labeled samples. The proposed method achieves an average accuracy of 85.9% for all subjects and the accuracy rates of single subject are between 60.5% and 99.3%. The average accuracy of learning new activity from another data is 84% and the accuracy rate of a single subject reached up to 100%.

Kwon et al. [138] proposed IMUTube integrating computer vision and signal processing techniques to convert human activity videos into virtual IMU data streams. These virtual IMU data streams are equivalent to accelerometer signals collected from various human body locations. The virtually generated IMU data can enhance the performance of different models on public HAR datasets. This work represents an integrated system that utilizes computer vision and signal processing for sensor-based on-body HAR that opens the door to further enhance the recognition accuracy based on a large dataset.

5 Transfer learning

Transfer learning refers to the learning paradigm where what has been learned in one setting and/or task is exploited to initiate and/or improve generalization in another setting and/or task [139]. It is the improvement of learning in a new task through the *transfer of knowledge* from a related task that has already been learned [140]. Transfer learning has been boosted by the recent revolution in deep learning, though it is not particularized to the latter. Deep neural computing has revolutionized *representation learning* which has made transfer learning readily available.

5.1 Fine tuning

The key motivation, especially in the context of deep learning, is the fact that most models which solve complex problems need much data, and getting vast amounts of labeled data for supervised models can be really difficult, considering the time and effort it takes to label data points. Hence, foundational models (i.e., models that have already been learned on massive amounts of data in capable entities such as big corporations) can be downloaded and *fine-tuned* for specific tasks with only limited amount of data with noticeable degree of success.

A rather recent survey on the progress of transfer learning for classification, regression, and clustering problems can be found in [141]. Lisa Torrey and Jude Shavlik in their book [140] described three possible

advantages to consider when employing transfer learning on a *target task* using a model pre-trained on a *source task*. These benefits are illustrated in Fig. 20, where two performance curves are shown: the green curve is the training performance using trained-from-scratch (end-to-end training) model on the target task whereas the red curve shows the performance of *fine-tuning* pre-trained source model to work on the target task.

1. *Higher start* The target skill starts higher than it otherwise would be. In some cases this can be considered as a satisfactory performance where fine-tuning may be unnecessary anymore.
2. *Higher slope* The rate of improvement of the target task during training (fine-tuning) of the source model is steeper than it otherwise would be. This means the convergence rate is higher.
3. *Higher asymptote* The convergence of the pretrained source model on the target task is better than it otherwise would be. This means it converges to a higher performance point.

On a more foundational level, the work done in [142] empirically investigates and quantifies the generality vs. specificity of neurons in each layer of a deep CNN. Their results indicate that *transferability* is negatively affected by two factors: (1) the specialization of higher layer neurons to their original source task at the expense of performance on the target task and (2) optimization difficulties related to splitting networks between co-adapted neurons (different hidden neurons having correlated behavior). They did an example network trained on ImageNet and demonstrated that either of these two issues may dominate, depending on whether the features are transferred from the bottom, middle, or top of the network. It is also derived, rather naturally, that transferability of features decreases as the distance between the base/source task and target task increases, but in any case transferability is better than random initialization (training-from-scratch). A final result is that initializing a network with transferred features from almost any level of layers can produce a boost to generalization, though depending on the layer the degree of effective transferability varies. Most of the research on transfer learning have been exclusively focused on visual data. Some recent work are directed towards timeseries specially inertial motion data for the purposes of activity recognition and gait analysis.

In the following, we survey some of the transfer learning work on timeseries data. Through the above-mentioned discussion we have already talked about several research works that in principle represent effective applications of transfer learning in the context of timeseries analysis. For example, Abdu-Aguye et al. [90] developed an approach to transfer learning for fixed and *variable-length* activity recognition inertial motion timeseries data. They do an end-to-end training using a CNN, see Fig. 17, then the convolutional filters are used as an *unsupervised feature extractor*. This portion of the network is then used to generate features for any dataset and activities; where such features can subsequently be fed to any classifier (in the paper, the classifier is a classical feedforward neural network).

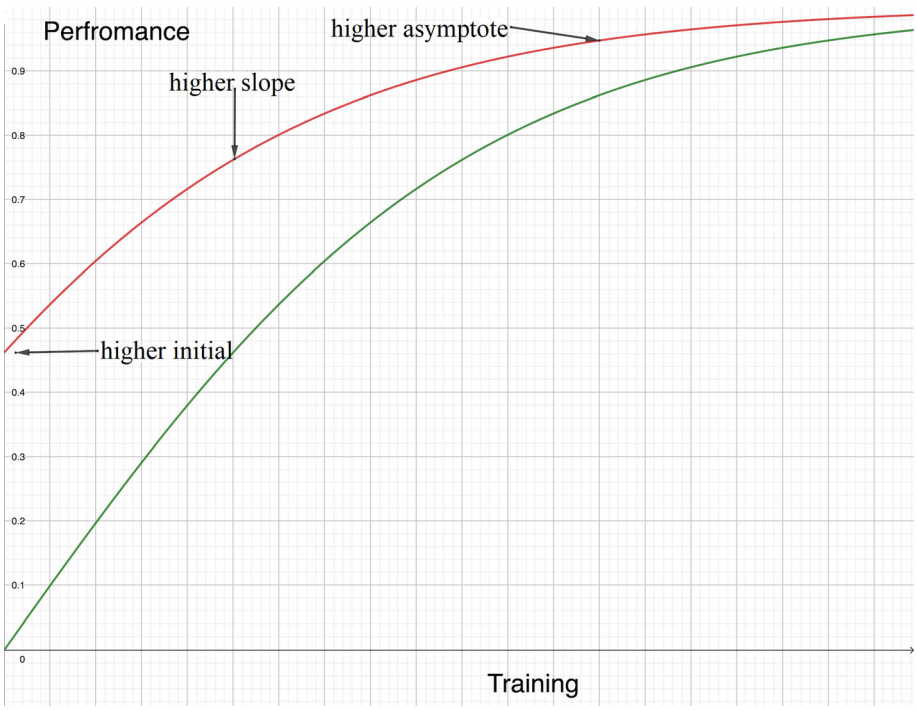


Fig. 20 Different ways of learning improvement through transfer learning

The authors used a mini-batch of size 128 for all the pooling parameters evaluations. The learning-rate number of training epochs in both self-testing and transfer learning was 35 epochs. The training and testing phases were repeated 15 times with different data samples were used each time for training and testing. In self-testing and transfer learning; training with 75% of data and testing with 25% of data.

The authors validated their approach on five different heterogeneous activity datasets, and a snapshot of their results is shown in Table 3. The classification accuracies obtained are shown together with their standard errors. (i.e., $accuracy \pm std.err$). The baseline are the diagonal elements indicating the self-testing scenarios (train and test on the same dataset). Whereas the off-diagonals show the transfer learning results (cross-testing): train on a dataset, fine-tune and test on another. Relative to the baseline diagonal accuracies, transfer learning shows very promising results, specially considering that the given datasets are very diverse regarding the devices used to collect the streaming data, the sampling rates, the subjects performing the activities, the locations of the sensors on the subject's body, etc. The results also indicate the stability and robustness of the transfer learning scenarios where the standard errors of the results are generally low, remaining below 3% in all the scenarios considered, and implying that transfer learning across the timeseries activity datasets is effective and comparable to training from scratch.

It can be observed that the HAD-AW and REALDISP datasets consistently have the best transfer learning (cross-testing) results. This can be attributed to the fact that they consist of a large and diverse range of activities, causing the convolutional filters to be relatively more discriminative than when trained on other smaller datasets. So, wider range of activities may play more crucial role in the effectiveness of transfer learning than having many and longer samples of a smaller set of activities. This conclusion is as well supported by observing that the HAPT dataset, having the smallest number of activities shows the poorest transfer learning results in all scenarios. *This makes a stronger case for the use of class-diverse datasets as source datasets for transfer learning.*

Now we look at the second important benefit of transfer learning, namely, *computational feasibility*. As transfer learning should save huge amount of computational resources as the target task requires just fine-tuning rather than training-from-scratch. The average times required for self-testing and transfer learning scenarios are shown in Table 4. The self-training scenarios take much longer times, due to the multitude of operations inherent in training the feature extraction part, namely the convolutional layers.

In transfer learning, the training is done on some dataset, then the classifier is fine-tuned on some other dataset. In this paper specifically, the authors train a CNN and used its convolutional filters as a feature extractor then trained a feedforward neural network classifier on the extracted features for other datasets. The transfer learning (cross-testing) scenarios take significantly less time (only for fine-tuning), yielding a 24.17× speedup compared to the diagonal baseline (end-to-end training) in the worst case (i.e., on the Daily Sports dataset) and a 52.31× speedup in the best case (i.e., on the REALDISP dataset). We can as well observe that the timings for the transfer-learning scenarios are fairly steady, regardless of the source dataset. This implies that the transfer learning testing times are purely dependent on the size of the target dataset.

The same authors have done another work in the same direction [92], where instead of using CNN as a feature extractor, they apply deep metric learning. *Metric learning* is based on representation learning of the input samples inside some embedding Euclidean space, such that, similar inputs (e.g., inertial timeseries from the same activity) are mapped to proximal representations in the embedding Euclidean space, and different inputs (e.g., inertial timeseries corresponding to different activities) are mapped to faraway representation vectors in the embedding space. Of course, the distance here is essentially the Euclidean distance. The "deep" qualifier in deep metric learning just refers to the use of the deep learning technology to generate such representations. The authors used a *deep Triplet network* to generate fixed-length descriptors for inertial timeseries samples for the purpose of activity classification.

The triplet network consists of 2 convolutional and 3 feed-forward layers. The authors involved between the convolutional and feedforward layers a 1-D Spatial Pyramid Pooling layer. Batch Normalization is used in between successive feedforward layers. The triplet loss function has a margin α parameter equals to 1. The triplet selection strategy was set to be the Random Negative Triplet Selection. The optimizer was the Stochastic Gradient Descent with a Nesterov Momentum value of 0.90. A 128-feature embedding vector was used as a feature vector for classification. For training, 75% of the selected datasets were utilized. After training, the training samples embeddings were used as exemplars for a 1-Nearest Neighbor classifier. This classifier is then used to obtain the embeddings classification accuracy of the unseen 25% of the dataset. Experimentation is carried out on the same five datasets as in Tables 3 and 4. Activity classification accuracies reached up to about 96% in the self-testing scenarios and up to about 91% in cross-testing without retraining.

Similar approach was taken up by Khaertdinov et al. [93] where they use deep metric learning equipped with

attention for HAR. Their goal is to develop a robust wearable-based HAR system that overcomes two difficulties: (1) inter-class similarities (different activities have similar inertial motion patterns) and (2) subject heterogeneity, where different subjects perform the same activity quite differently. The hyperparameter tuning was done on the datasets validation sections. The models parameters were optimized by the Stochastic Gradient Descent with Adaptive Moment Estimation (ADAM) using the default parameters ($\epsilon = 10^{-8}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$). The models were trained for period of 100 epochs. The learning rate was doubly decreased after every 10 epochs without performance improvement. Training with Hierarchical Triplet Loss (HTL) needs pre-training with the original triplet loss. Thus, the best models on the validation sets were fine-tuned using HTL for 10 extra epochs. For all models (including the LSTM model trained with the original triplet loss), the authors made of the LSTM hidden state vectors size equals to 512. For all the triplet networks, the authors fixed the mini-batch size near to 128.

Hu et al. [143] did work aiming at re-purposing activity instances collected in some activity domain for training a classifier in some other activity domain. The authors call these samples "pseudo training data" in the target domain. Contrasted with the previous work, which is *feature-based*, this kind of transfer learning is *instance-based*. They achieve their goal by developing an algorithm that utilizes pre-existing web data describing the two activity sets. Using such descriptions a similarity framework is built within which activity samples in the source domain are mapped to activity samples in the target domain with some confidence value.

The authors are omitting sequential information in the dataset, Thus, the evaluation criteria is just to calculate the accuracy on the test dataset; that is the correctly predicted activities over the total activities number in all time periods. Using the activity names (e.g., preparing breakfast, cleaning, etc.) as queries submitted to Google and retrieving the top K results; K is a parameter tuned to clarify the relationship between the algorithm accuracy and the Web pages number retrieved for each query. Using that approach the authors managed to get significant performance increase in activity recognition in the target domain as compared to the baseline case where no knowledge transfer is done. In experimentation over real world datasets, their algorithm achieved about 60% accuracy with no or very few training data in the target domain.

Khan et al. [144] proposed an instance-based transfer learning framework. The work is quite motivated as they emphasize the need for transfer learning. Their justification can be articulated in the following points:

1. HAR systems are typically based on a limited number of activities that are present in the training and testing splits of the underlying dataset. However, in practical scenarios these limited set of activities need to be extended (concept drift).
2. In practical scenarios the data distributions can be radically different from the distributions inferred during the training phase of the HAR system (data drift).
3. There is shortage in labeled data samples, and abundance of unlabeled samples. Transfer learning can help in that direction of *self-supervised* learning.
4. Training and testing environment and setup are typically quite different from the corresponding deployments.

Many more can be said about the challenges in inertial motion-based HAR systems. Some detailed discussion is given regarding the heterogeneities in HAR systems in Sect. 1. See as well Fig. 20 and the associated discussion at the beginning of Sect. 5 about the learning benefits gained from transfer learning.

In the face of such challenges, Khan et al. [144] proposed a framework, called *TransAct*, by augmenting the Instance-based Transfer Boost algorithm. They apply their method on three publicly available datasets: MHealth [95], Daily and Sports Activities [57], and SBHAR [42] (detailed description of these datasets can be found in Table 15). The authors experimented with these datasets in such a way to ascertain their transfer learning agenda. They do cross-testing spanning different environments, and accommodating new sets of activities and changes in the data distribution by using a combination of clustering with anomaly detection.

The accelerometer data was segmented by sliding window with frames of 128 samples with 50% overlap. The frame length was the same across all the datasets. The authors removed noise using low-pass filter. Using FFT on the accelerometer data, most of the high-energy frequency components were lying in-between 0 and 20 Hz. The authors extracted both time domain and frequency domain features. Time domain features (e.g., mean, standard deviation, etc.) and frequency domain features (e.g., energy, entropy, etc.) are calculated using FFT on each frame. The experimental results show that the *TransAct* model achieves on average 81% activity detection accuracy.

Khan et al. [145] proposed a framework, called *UnTran* that aims at bridging the gap between a learned HAR model in some source domain and its adaptability or deployment in a different activity target domain (device heterogeneities, sensing biasness, inherent variabilities in subjects' behaviors and actuations). *UnTran* is based on

transferring two layers of an autoencoder pretrained on samples from the source domain to the target domain, where such two layers are used to generate common feature space for the two domains. Specifically, the authors trained a Deep Sparse Autoencoder (DSAE) on pre-extracted feature vectors and subsequently use a part of the pre-trained encoder as a generic feature extractor. Three SVM classifiers are trained across three different configurations: (1) training on features extracted by the DSAE over the source domain samples, (2) training on features extracted from DSAE over the target samples, and finally, (3) training on low-level features extracted from the target samples. The decisions from the three classifiers are then fused to induce a final decision. The authors validated their work on three datasets: Opportunity [146], Daily and Sports Activities [57], and WISDM Actitracker [106] (detailed description of these datasets can be found in tab 15).

The authors segmented the accelerometer data into frames of 128 samples with 50% overlap between frames. Low-pass median filter was applied to cancel noise. Statistical time-domain and frequency-domain features were extracted. These features were fed to the classifier in batches each with a batch size of 32. The frame length and batch size were the same across all datasets and experiments. Transfer learning baseline methods were implemented; Transfer Component Analysis (TCA) [147] in Python and Joint Distribution Adaptation (JDA) [148] in MATLAB. The proposed Deep Sparse Autoencoder (DSAE) is comprised of 4 layers. A softmax layer was used to encode the classification labels in the source domain. The first two layers of the source tuned network were used to build the proposed classifier.

The experiments show that *UnTran* achieves about 75% F1-score in recognizing unseen new activities using only 10% percent of the target samples. The system also maintains an F1-score of %98 for recognition of seen activities in presence of only 2–3% of labeled activity samples.

Motivated by almost the same reasons, Chikhaoui et al. [149] explored the development of a transfer learning framework based on CNN. Learned feature representations in one activity dataset are used to analyze activities in another dataset that is characterized by different data collection hardware and software setups, and even changes in the sensing modalities. A CNN classifier is pretrained and the learned weights are transferred to a new similarly structured network in the target domain. The latter network is fine-tuned using samples from the target dataset without making any changes or updates to the transferred weights. The authors validated their work over three benchmark datasets: MobiAct [150], RealWorld Human Activity Recognition (RWHAR) [151], and Mobile Sensing Heterogeneities for Activity Recognition [152]. Hence, they tested the effectiveness of their approach across different users, device placements, and sensors modalities.

The accelerometer data was segmented into non-overlapping 4 s segments. The authors trained the source model using 100 epochs and the target models were fine-tuned with 50 epochs. The CNN models were implemented using the Tensorflow library. The authors obtained F1-scores up to about 0.94 in some of the evaluated scenarios.

Another interesting work with a different spirit is that done by Abdu-Aguye et al. [25]; we briefly discussed this work above. The idea can be conceptualized as generating *"virtual IMU sensors"*. Given inertial motion data streamed from a real IMU sensor mounted on some source body location, these timeseries are then mapped to their correspondence on a target body location. So, these latter samples can be considered as streaming inertial motion data from a *virtual IMU sensor* mounted on the target body location. Independent of the performed activity the authors developed a learning architecture that maps an inertial sample from the source body location to the target body location. It is emphasized that this is done regardless of the activity, it is meant to capture how dynamics in general are transferred and correlated across different body locations. In order to empirically test their work they developed a rigorous evaluation strategy as shown in Fig. 21. Given a source body location and a target body location the dataset inertial samples are divided into 3 groups: 50%, 25%, and 25%. The mapping between the source and target locations is called "roaming". Three sets of evaluations are performed. The first kind of evaluation is aimed at establishing a baseline as a lower bound for the predictive performance: a classifier is trained on 25% of the target samples (1B in Fig. 21), then testing on 25% of the source samples (0C) using this classifier. The second evaluation is based on training a roaming model on the 50% source and target samples (denoted as 0A and 1A Fig. 21), the input to the roaming model is a source sample and the output is a target sample. Then, roam 25% of the source samples (0C) using the trained model. The classifier developed in the first evaluation is then used to evaluate the roamed samples, that is, the roamed samples form the testing data. This evaluation helps deriving the performance gains obtained from using the roaming model. The third and final set of evaluations is done to obtain an upper bound of the predictive performance, where the trained classifier, from the first set of evaluations, is used to evaluate genuine target samples (1C). As designed the predictive performance of the roaming model (the second set of experiments) lies between the lower and upper bounds obtained by the first and third sets of evaluations. The authors used the REALDISP [13] data set for their empirical study, where multiple sensors stream from different body locations at the

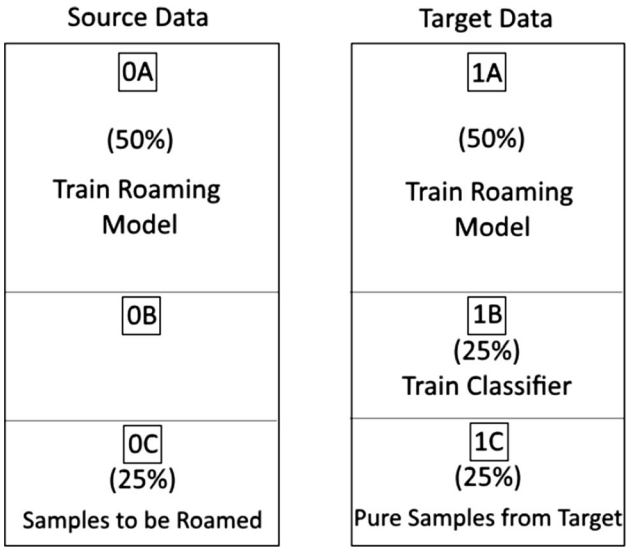


Fig. 21 Data assignment for experimental and evaluation methodology [25]

same time; see Table 15. A recurrent deep model, based on Bi-directional LSTM is used for training the roaming map, see Fig. 8.

Table 11 shows a sample of the results obtained through roaming in terms of average accuracy and F1-score of activity recognition. The left column shows the activities categorized into four groups: activities involving the whole body (‡), trunk activities (‖), the upper extremities (†), and the lower extremities (^). The source location is the right lower arm (RLA) and two target locations: the right upper arm (RUA) and back (BACK). Except for activities involving the lower extremities the RUA upper bound (the UB column) shows good F1-metrics. This can be trivially explained by the fact that such activities have more dynamics captured by the RUA as compared to those involving only the lower extremities. In addition, the lower bounds (the LB column in the table) indicate that the upper and lower parts of the arm generate similar dynamics; this may be attributed to their proximity and connectedness. The targeted proposed roaming (the "Proposed" column in the table) then show significant improvement over the lower bound indicating the effectiveness of the roaming model. Turning into the back location we see the lower bounds results are much worse. This means that the classifier is unable to detect the proper dynamics of the back from the RLA. This emphasizes the need for transformation, 'roaming' using the authors' terminology, of the inertial motion signals of the RLA to the back before making any useful inference. As shown in the table the transformation is rather effective in many of the cases, specially when comparing to the upper bounds.

5.2 Prompting

With the rise of pre-trained models, fine-tuning these models to target domains is a crucial task. Thus, parameter efficient transfer learning of large models is very important. In [153], the authors proposed black-box visual prompting (Black-VIP) that adapts the pre-trained models efficiently without knowing the model architectures and parameters. Black-VIP consists of two components; coordinator and simultaneous perturbation stochastic approximation with gradient correction (SPSA-GC). The coordinator designs input-dependent image-shaped visual prompts that enhances the few-shot adapting on distribution and location shift. SPSA-GC estimates efficiently the target model gradient to update the coordinator. Experiments on 16 datasets showed that BlackVIP allows robust adaptation to different domains with no access to the pre-trained models parameters using lowest memory requirements. Such prompting models can help in communicating with the feedback of a virtual coach for human activity recognition and assessment [154].

In [155], the authors proposed "inverse prompting" to control better the generation of text. The main idea of inverse prompting is using the generated text in order to inversely predict the prompt that improves the relevance between the prompt and the generated text. The authors pre-trained a large-scale language model for a systematic study using human evaluation on tasks of open-domain poem generation and question answering. The results showed that the proposed approach outperforms the baseline models and the generation quality was near to the human performance in some tasks.

In [156], the authors proposed a Prompt approach for Text Generation (PTG) in a transferable setting. Initially, PTG learns some source prompts for different source generation tasks and then these prompts are transferred for target generation tasks. The authors designed an adaptive attention approach to calculate the target prompts that includes task-level and instance-level information. For each data instance, PTG learns a specific target prompt attached to highly relevant source prompts. The results showed that PTG achieved better results than fine-tuning approaches. The authors released the proposed source prompts as open source where researchers are able to reuse in order to enhance new text generation tasks.

In [157], the authors presented Multitask Prompt Tuning (MPT) that initially learns a single transferable prompt by transferring knowledge from different task-specific source prompts. The authors then learned multiplicative low rank changes to this prompt to adapt efficiently to every downstream target case. Results on 23 NLP datasets showed that the proposed method outperforms the state-of-

Table 11 F1-measures for roaming experiments for right lower arm (RLA) to right upper arm (RUA) and back (BACK) [25]

Source location	RLA					
Target location	RUA			BACK		
Method	LB	Proposed	UB	LB	Proposed	UB
Average classification accuracy (%):	21.21	66.05	73.48	4.66	44.18	63.**35**
Walking‡	0.55	0.71	0.83	0.00	0.63	0.78
Jogging‡	0.04	0.72	0.77	0.00	0.54	0.70
Running‡	0.02	0.70	0.79	0.01	0.63	0.85
Jump up‡	0.28	0.52	0.59	0.02	0.38	0.40
Jump front and back‡	0.36	0.56	0.64	0.16	0.41	0.68
Jump sideways‡	0.11	0.41	0.50	0.11	0.44	0.65
Jump legs/arms open/closed‡	0.21	0.86	0.92	0.04	0.35	0.54
Jump rope‡	0.24	0.57	0.60	0.03	0.24	0.35
Rowing‡	0.07	0.64	0.66	0.01	0.55	0.64
Elliptic bike‡	0.00	0.71	0.63	0.00	0.46	0.56
Cycling‡	0.05	0.57	0.78	0.00	0.11	0.47
Trunk twist (arms out)‖	0.41	0.83	0.85	0.02	0.55	0.59
Trunk Twist (elbows bent)‖	0.00	0.94	0.94	0.04	0.51	0.63
Waist bent forward‖	0.16	0.56	0.75	0.01	0.69	0.92
Waist rotation‖	0.02	0.65	0.76	0.05	0.45	0.77
Waist bends (reach foot w/ opposite hand)‖	0.27	0.78	0.92	0.00	0.67	0.87
Reach heel backwards‖	0.46	0.52	0.67	0.20	0.29	0.56
Lateral bend‖	0.14	0.51	0.54	0.11	0.42	0.62
Lateral bend arm up‖	0.17	0.41	0.49	0.05	0.32	0.57
Repetitive forward stretching‖	0.15	0.68	0.79	0.00	0.63	0.90
Upper trunk and lower body opposite twist‖	0.13	0.42	0.61	0.01	0.10	0.51
Arms lateral elevation†	0.12	0.70	0.59	0.00	0.10	0.23
Arms frontal elevation†	0.07	0.69	0.72	0.00	0.14	0.25
Frontal hand claps†	0.38	0.68	0.78	0.00	0.25	0.40
Arms frontal crossing†	0.29	0.83	0.84	0.00	0.17	0.49
Shoulders high-amplitude rotation†	0.72	0.89	0.91	0.01	0.13	0.33
Shoulders low-amplitude rotation†	0.22	0.87	0.83	0.01	0.23	0.29
Arms inner rotation†	0.06	0.76	0.72	0.00	0.11	0.60
Knees (alternately) to breast^	0.17	0.71	0.86	0.14	0.63	0.88
Heels (alternately) to backside^	0.35	0.73	0.73	0.23	0.59	0.80
Knees bending (crouching)^	0.31	0.31	0.35	0.24	0.32	0.60
Knees (alternately) bent forward^	0.18	0.43	0.59	0.02	0.28	0.54
Rotation on knees^	0.01	0.69	0.69	0.00	0.56	0.74

the-art approaches including the full fine-tuning baseline method in some tasks though tuning only 0.035% of task-specific parameters.

6 Application domains

In this section we demonstrate some of the case studies for which inertial motion sensors are effectively used for different kinds and tracks of human actions.

6.1 Biometrics

Human biometrics are a set of measurements of bodily characteristics that are to be used for authentication and/or identification. The most common biometrics include: fingerprints, facial, voice, iris, palm or finger vein patterns. However, in our context the biometrics are based on the execution pattern of an action or activity, through the use of streaming IMU motion signals. Recognition of *human biometrics* has been widely studied in the recent years due to its significance and use in many applications such as security and surveillance, speech analysis [158, 159], recommendation systems [160], and healthcare applications [161].

The wide and varying types of biometrics dictate the use of various kinds of data modalities streaming from different kinds of sensors. The typical sensors used are the visual, audio, and inertial motion data. Here, we naturally focus on the latter, specially that its applicability is more recent with the wide spread use of wearables with embedded IMU units.

To the best of our knowledge IMU-based biometrics systems have been exclusively directed towards three characteristics: age, gender, and gait. The next subsection is dedicated to gait analysis, and here we consider age and gender. Mostafa et al. [19] proposed a method for gender recognition from inertial data based on wavelet transform (refer to Sect. 3.2 for details about feature extraction using Fourier transforms and its variants). The motion data are collected during walking activity (gait streaming data). They co-collect data for that purpose, which is the EJUST-GINR-1 [19, 20], which contain samples from smartwatches in addition to specialized IMU units placed at 8 different parts of the human body; see Fig. 9. Random forest and CNN are used for the binary classification: male/female; with input features the wavelet transform of the 6-axial timeseries: 3-axis accelerometer and 3-axis gyroscope angular velocity. The authors studied which sensor location(s) on the human body is(are) optimal in terms of distinguishing males from females during the act of walking. They found, rather expected, that sensors placed on the legs and waist are best in recognizing the gender during walking. They reached accuracy of about 97% using the sensor placed on the left cube. The authors did experiments as well with the OU-ISIR Gait dataset [33], however, the performance was rather modest about 75% (using the CNN).

Using the OU-ISIR dataset [33] a competition for gender recognition and age estimation was conducted [43] based on IMU signals extracted during gait activity. Most of the participated teams got relatively low accuracy in gender classification. On the other hand, results on age estimation were much better. The best results of this competition are presented by Garofalo et al. [162], which reached accuracy of about 76% for gender recognition and a mean absolute error of about 5.39 for age regression by using orientation independent AE-GDI representation along with a CNN.

Riaz et al. [163] proposed a system for gender recognition, age and height estimation based on IMU signals streamed from on-body sensors during gait action. Age estimation is formulated as a classification problem by splitting the age into three ranges: (1) less than 40 years, (2) between 40 and 50, and (3) older than 50 years. Height estimation is formulated as a classification problem by splitting the height into three ranges: (1) less than or equal 170 cm, (2) between 170 and 180, and (3) greater than or equal 180 cm. The authors used random forest as a classifier with two kinds of validation strategy: tenfold cross validation and the more difficult subject-wise cross-validation. They reached an accuracy about 93% for gender recognition using the sensor on the chest, about 89% for age classification using chest and lower back sensors, and about 89% for height classification using the chest sensor.

Jain et al. [164] did gender recognition, from gait action, using accelerometer and gyroscope signals streamed from IMU sensors embedded onto a smartphone. The authors used multi-level local pattern and local binary pattern as feature extractors. The extracted features are then fed to an SVM and aggregate bootstrapping (bagging). To evaluate these models, 252 gait signals collected from 42 subjects were used. They achieved an overall accuracy of about 77% using multi-level local pattern as the extracted features along with bagging.

Mostafa et al. [32] proposed a system, called BioDeep, for gender and age analysis. The author employ a simple, yet effective as shown above in other works, feature extraction technique which is based on the autocorrelation function. This descriptor is then fed to two decoupled CNN networks, one for gender recognition and the other for age estimation. Age estimation here is formulated as a regression problem. As a baseline they also used random forest justified by the fact that they test the effectivity of deep neural networks. However, there is some confusion in the paper as they claim to test the effectivity of deep learning, rather, they test the effectivity of deep neural networks in particular, as random forest is a deep learning strategy as well [165]. The main contribution in the paper can be considered as the extensive empirical studies they performed on a varying list of datasets. To the best of our knowledge, this might be the most comprehensive study from that perspective. They perform self-testing as well as cross-testing (transfer learning) across these datasets. The proposed system architecture is shown in Fig. 13. The authors validated their work using several datasets and

different on-body sensors including: EJUST-GINR-1 (right cube RC sensor), EJUST-GINR-1 (Waist sensor) [19, 20], HuGaDB (Right Shin RS) [22], and GEDS (Tibialis anterior Right TaR) [34].

Tables 12 and 13 show a sample of the results produced in this paper both for gender recognition and age estimation. The diagonal elements show the self-testing scenarios, that is, train and test on the same dataset, which also play the role of a baseline when considering the off-diagonal elements which represent the cross-testing results, that is, train on a dataset and testing on a completely different one. The performance metric for gender recognition is the overall accuracy, whereas it is the mean absolute error for age estimation. The results indicate that the proposed method seems very effective in both self-testing and cross-testing as well, where the latter is more significant as it represents a sort of transfer learning that can be used to overcome the heterogeneity problem in inertial-based systems (see the discussion in Sect. 1 for the sources and consequences of this problem). For gender recognition, the results are transferable to within 1–2%; for age estimation the results are transferable to within 2–3 times increase in the mean absolute error. The authors claim that transfer learning provided 20–30× speedup in the training time compared to end-to-end training on the same dataset.

Riaz et al. [166] estimate the age based on the IMU analysis of human walk. The authors recorded 6-dim accelerations and angular velocities of 86 subjects when they perform standard gait activities using IMUs mounted on chest. The recorded signals were segmented into single steps; for each step, a total of 50 spatio-spectral features were computed from the 6 dimensional components. In order to estimate the subject age, three ML classifiers were trained: RFs, SVMs, and MLP. Two cross validation techniques were utilized: tenfold and subject-wise cross-validation. A random forest regressor trained and validated on hybrid data achieved an average RMSE (root mean squared error) of 3.32 years and a MAE (mean absolute error) of 1.75 years with tenfold cross validation and average RMSE of 8.22 years with subject-wise cross validation. Since the subjects belong to two demographical regions (Europe and South Asia), this gives broader empirical confirmation that age can be estimated based on human gait.

Khabir et al. [167] estimated gender and age from IMU-based gait dataset that is part of the Osaka University-ISIR Gait Database [43]. The authors showed that a SVM model had the best accuracy for classifying gender while Decision Trees had the highest variance score (R^2 value) for age prediction.

Rasnayaka et al. [168] studied the personal characteristics collected from on-body IMU sensors and the sensor location effect on the user's privacy. By analyzing the sensor locations with respect to privacy and service, the authors recommended locations have maintained the functionality (like the biometric authentication) while reduced the privacy invasions. The authors collected multi-stream dataset from 3 on-body IMU sensors consisting of 6 locations for 6 actions with different physical, personal, and socio-economic characteristics from 53 subjects. An opinion survey was conducted about the relative importance of each characteristic from 566 persons. Using these datasets the authors showed that the gait data can tell much personal information that may be considered a privacy issue. The opinion survey gave a ranking of the physical attributes based on the recognized importance. Using a privacy vulnerability index, the sensors located in the front pocket/wrist were more privacy invasive compared to the back-pocket/bag sensors that were less privacy invasive without a big loss of the functionality (e.g., the biometric authentication).

6.1.1 Gait analysis

Gait is the manner of human walk, and *it is a biometric trait that is unique to each individual* [169]. In the computational domain, gait analysis has importance, particularly, in *person authentication* and *identification*. Person authentication is concerned with the problem of identifying if a new gait instance belongs to a legitimate user or is an imposter [170]. This is typically done by searching for the new pattern in a list of patterns of the legitimate users [171]. In contrast, for person identification/recognition, the task is to classify a new gait instance to belong to one of a set of predefined users [43]. The embedded inertial motion sensors in the commodity wearable devices provide real-time streaming of linear acceleration and angular velocity, in addition to magnetometer readings, which together can

Table 12 Self- and cross-testing results of gender recognition (accuracy) [32]

Training dataset	Testing dataset			
	EJUST-GINR-1(RC)	HuGaDB(RS)	GEDS(TaR)	OU-ISIR(Waist)
EJUST-GINR-1(RC)	97.05%	98.13%	96.56%	NA
HuGaDB(RS)	95.58%	99.37%	96.2%	NA
GEDS(TaR)	95.9%	98.03%	97.47%	NA
EJUST-GINR-1(Waist)	NA	NA	NA	67.93%

Table 13 Self- and cross-testing results of age estimation (mean absolute error) [32]

Training dataset	Testing dataset			
	EJUST-GINR-1(RC)	HuGaDB(RS)	GEDS(TaR)	OU-ISIR(Waist)
EJUST-GINR-1(RC)	0.883	2.0	2.5982	NA
HuGaDB(RS)	2.5685	0.835	3.6507	NA
GEDS(TaR)	2.232	2.19	1.963	NA
EJUST-GINR-1(Waist)	NA	NA	NA	5.1209

fully capture the motion dynamics of the owner's motion, or at least the motion dynamics of the body part (for example, wrist) to which the device is attached. The main advantage here is that such process occurs implicitly and does not require the active participation of the human subject [172–175]. This is in contrast to fingerprint authentication where the person has to take an action for authentication or even to face recognition which can be easily fooled. *Person identification* using motion data streaming from wearables turns out to be important in applications, for example, when different people share the same wearable device and it is desirable to identify who is using the device at a given time while walking.

The work done in [20] tries to answer two questions (though the answers seem obvious, but it is more of an engineering verification): (1) what is(are) the body location(s) that is(are) most effective for person identification through streaming from an IMU device attached to such part(s) and (2) if we have the freedom of mounting more IMU devices to different body parts, then which combination is the most effective? The authors apply their work on the EJUST-GINR-1 [19, 20] dataset. It is a gait dataset with 20 subjects, 10 males and 10 females. See Table 15 for a full description of the dataset; and see Fig. 9 for a depiction of the sensors locations over the different body parts. Two of these sensors are smartwatches worn on both wrists. The authors use all the 3 axes of the linear acceleration as well as the 3 axes of the angular motion as streaming sources. Then, they use derivatives of these timeseries, specifically, the autocorrelation function, to derive 6-axial fixed size much smaller signals. These are fed to 3 different classifiers, namely, random forest, support vector machine, and a convolution neural network. The overall schema of the system is shown in Fig. 19. Detailed performance metrics are given for every sensor on the body and for paired combination of sensors for the three classifiers. Generally, the results seem good, and, as expected, the best results are obtained from sensors mounted on the lower part of the body as they are closer to the motion source. However, the sensor on the back of the neck achieves next best performance to those IMUs mounted on the lower body parts. The authors hypothesize that the "gait patterns are transmitted almost intact through the backbone up to this location", however, they do not provide any further evidence of references to this claim. The main drawback of this work is that the number of subjects are small (total of 20), in addition to the lack of diversity in their age (they are all early 20s), though they are balanced regarding gender.

An extension of this work was done by Adel et al. [176], by using a more advanced modeling technique, namely, Siamese networks, and validating the work over multiple richer datasets. In a sense, this work brings the long standing work of visual-based gait recognition using Siamese networks into the realm of inertial motion data. As expected these networks are used for automatic feature extraction. The proposed scheme is used for person authentication, where the network is trained, using raw inertial data, with only few samples per authorized subject. Once trained, the network can then be used to authenticate any given user using only one gait sample of that user. The work is validated using three publicly available datasets: OU-ISIR [33] (large-scale dataset with 744 subjects), EJUST-GINR-1 [19, 20] (small scale dataset with 20 subjects only), and MMUISD [23] (medium scale dataset with 120 subjects). The main performance metric adopted in this paper is the equal error rate (EER). The EER is the point on the ROC (receiver operating characteristic) curve where the false acceptance/positive rate equals the false rejection/negative rate. It is one of the typical metrics used to measure the performance of person identification and authentication systems. Accuracy is reported as well. The overall framework is shown in Fig. 22. Their model achieved 3.42% EER on OU-ISIR (it is the world's largest inertial gait dataset) after training on 592 subjects and testing on other 78 subjects; on average each subject provided an average of 5.9 s of gait data. The best EER on the MMUISD was 6.26%, and 5.13% on the EJUST-GINR-1. The latter dataset has the fewest number of subjects (only 20), but the largest gait duration of each subject. So, we can conclude from these results that *larger number of subjects is more effective than the gait longevity of the subjects*. An interesting work done by the authors is the use of transfer learning to train on subjects from a particular dataset and authenticate subjects from another. The hardware and software setup across the different datasets are quite different (for example, the type of the sensor and its location on the body); see Table 15. Hence, this task is quite

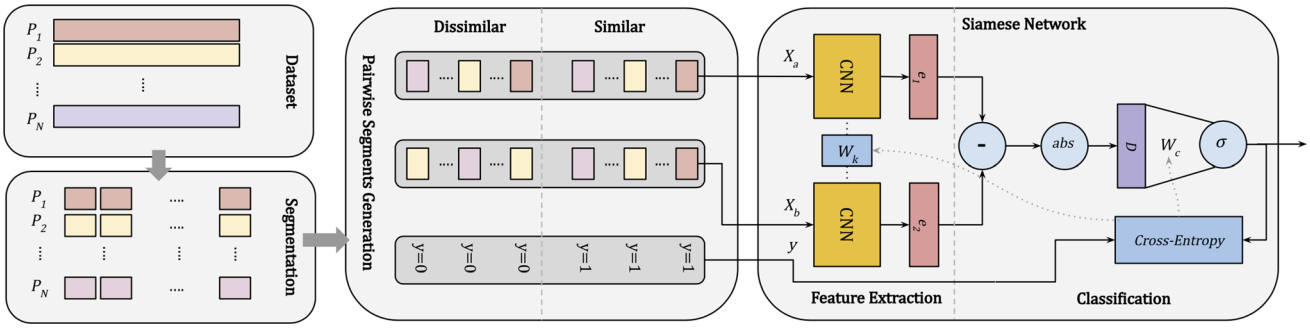


Fig. 22 Framework based on Siamese networks for person authentication based on inertial gait streaming [176]

difficult. The base case of performance is, of course, when the source and target datasets are the same. The trained Siamese network is essentially used to generate embeddings and, then, difference vectors, for the target datasets. Applying the same reasoning as above using the OU-ISIR as the source dataset produces the best transfer results over the other datasets.

6.2 Virtual coaching

In this section, we survey the activity recognition works mainly focusing on indoor exercises (push-ups, squats, etc.). In the context of athletics and sporting, inertial motion sensors are typically made up of accelerometers to measure force and acceleration, a gyroscope to give an indication of rotation (angular velocity and angular displacement), and a magnetometer to measure body orientation. Each of these sensors collects data across three perpendicular axes and can capture athlete's movement in minute detail. These are used to continuously stream data that are typically processed by state-of-the-art machine learning, data modeling, and inference techniques that induce much higher level useful information from such data for workout and exercising actions. For example, they can be used to differentiate between a jump to the right or a jump to the left, a walk or a run, a walk on sand, solid land, or grass, etc. Other uses include measuring bowling in cricket, jumping in basketball, or kicking a ball in soccer, fencing, etc. One of the most popular uses of inertial sensors is in trying to quantify athlete readiness and fatigue in the field. For example, there are protocols that are often included in athlete's warm up while wearing sensors. Wearables, especially IMUs, can be used to detect changes in acceleration or direction of acceleration which may change with injury and/or fatigue. Each person has an *individual motion signature*, so coaches, trainers, and sports scientists can compare an athlete to her typical *normal* self. IMUs can provide good insights into the exercising and/or playing performance and can aid in providing assessment and feedback to athletes recovering from injury.

Generally, IMUs can be placed anywhere on the human body. However, depending on the exercise or workout, some places are more proper to capture the dynamics of the motion of that exercise. For example, IMUs can be placed at the ankle for running sports, at the waist to detect jumping, at the wrists for sports such as tennis and fencing, and in between the scapula to approximate rough body movements. Current IMU technologies have increased dramatically, for example, sampling rates have exceeded 1 kHz, which allows for thorough and detailed analysis of high speed movements. Also, dynamic ranges have increased to about 200 g (for force and acceleration measurements), so we can now measure movements such as high velocity cutting, acceleration and deceleration, and jumping.

Optimal human movement is an important goal for coaches, trainers, therapists, and athletes in most sports. Skeletal muscles generate the force needed for these motions. Consequently, proper timing of skeletal muscle activation is important to improve movement quality and achieve better sports performance. However, injury, fatigue, and/or poor training can lead to sub-optimal timing and coordination of skeletal muscle activations [177, 178]. Consequently, this can lead to severe decrease of exercising performance and increase the risk of injury. Therefore, the assessment of muscle activation patterns provides a picture of the overall health and effectiveness of the musculoskeletal system [179]. It allows athletes to reduce muscle injury or imbalance and increase the effectiveness of training exercises.

Different feedback systems are used in the field of *sports* and *rehabilitation*. The use of such systems has shown some success by providing a positive effect on training control [180, 181]. In addition, the feedback system has the advantage of providing early warnings and precautions that can help to reduce the risk of injury [182] and to improve the performance during different movement tasks (e.g.,

balance) of elderly people [183]. Feedback can be given in several ways including visual, auditory, haptic, textual, or various combinations of these.

In [31], the authors present a multi-modal deep learning method utilizing many data sources for activity segmentation, exercise recognition, in addition to counting the number of repetitions. The authors introduced the *MM-Fit*, which is a multi-modal dataset containing a collection of IMUs extracted from multiple devices: smartphones, smartwatches, and earbuds. These devices were worn by subjects performing full-body workouts and are time-synchronized with multi-viewpoint RGB-D video, and pose estimates in 2D and 3D. They have developed a very rigorous architecture that is hierarchical in nature, where at the bottom level, autoencoders are used to learn representations of the individual modality signals. Then, at later layers incremental fusing of the representations of different modalities is done in order to learn more abstract representations (embeddings) of the learned lower level representations. Three main tasks are addressed by this work: activity segmentation, exercise recognition, and counting the number of repetitions of a given exercise. Totally, 21 full-body workout sessions are recorded; totalling 809 min across 10 participants; workout lengths range 27–67 min with average duration 39 min. Workout exercises monitored in this dataset include: squats, lunges (with dumbbells), bicep curls (alternating arms), sit-ups, push-ups, sitting overhead dumbbell triceps extensions, standing dumbbell rows, jumping jacks, sitting dumbbell shoulder press, and dumbbell lateral shoulder raises. Table 15 gives more detailed description of the dataset. For exercise recognition the authors did extensive evaluation using each device separately and individual modalities within device, in addition to different fusion of all these. For example, taking IMUs, the authors experiment with accelerometer only, gyroscope only, and a combination of both. The best recognition accuracy is obtained from the fusion of all sensors' modalities reaching more than 99%. On the other hand, the inertial sensors of the earbud and the smartphone put in the right trouser pocket gave the lowest accuracy around 80%. Similar fine extensive evaluation has been done regarding the performance of counting repetitions of each exercise across each device. The performance metrics used were mean absolute repetition counting error and the percentage of exercise sets for which the predicted repetition count is exact, within 1, or within 2 of the ground truth repetition count. From the perspective of exercise type, squats, lunges, and lateral raises gave the lowest mean absolute error regarding their repetition counting. On the other hand, biceps curls and push-ups gave the worst mean absolute error. These results are consistent across all devices, sensors, and modalities. From the device perspective, on average across all exercises, the smartwatch on the left hand gave the lowest mean absolute repetition counting error.

6.3 Healthcare

Accurate and reliable timely quantification of human psychological, physical and physiological state using wearable sensing devices is paramount in health monitoring and safety surveillance. This in accordance necessitates the measurement of relevant signals from various on-body spots. A comprehensive review on the use of wearables in the medical sector can be found in [184]. All the capabilities of smartphones, including on-board embedded IMU sensors, can be used to collect *behavioral data* unobtrusively (without any intervention or disturbance to the subject), in situ, as it unfolds in the course of daily life [185]. Harari et al. [185] investigate the use of continuously streaming data of a person's social interactions, daily activities, and mobility patterns for *psychological* research and *profiling* and for discussing the ongoing methodological and ethical challenges associated with research in this domain. The authors boldly claim that smartphone, and generally wearables, sensing would lead the psychology on the road to fully realizing its promise as a truly behavioral science.

The work done by Baghdadi et al. [186] aims at developing a general framework for a monitoring system for the prediction of critical physical states prior to any potential safety and health incident. They used Kalman filter to estimate the hip acceleration and trunk posture from a data streamed from a single IMU mounted at the ankle. Such data is combined with apriori information from Fourier series approximation to relate body kinematics considering the periodic nature of gait. As a first application they built an SVM classifier for the continuous monitoring of *fatigue* by classifying the feature pattern of the kinematic profile categorizing into 'fatigue' versus 'non-fatigue' states. This provides for minimally intrusive, inexpensive approach for fatigue detection in the workplace. The authors examined a couple of research questions: (1) can a single IMU unit placed at the right ankle be sufficient to capture the subtle changes in gait due to fatigue? and (2) can a computationally efficient classifier be used to distinguish between fatigued and non-fatigued states? The authors did experiments using SVM as a classifier and the results affirmed positively these questions. They managed to reach an accuracy of 90% in predicting the fatigue states for the 20 recruited subjects. This study found that the fatigue, due to a simulated manufacturing task of manual material handling, results in changes in gait kinematics. So, in a more abstract sense, this is saying that inertial motion data are able to distinguish and capture the *fatigue* state. From the feature extraction/

selection perspective, not all features of gait kinematics have the same performance in fatigue prediction; foot acceleration and position trajectories are among the high performing ones. This line of research is crucial for studies in ergonomics. The authors projected that their research studies propose a framework for *safety surveillance* and *fatigue monitoring* to prevent safety incidents utilizing a minimal set of sensors and data that can be extended to other movement monitoring applications in healthcare through the ordinary smart wearable devices.

IMU devices typically have local storage that is large enough, so data can be recorded directly to the device, hence, sampling rates can be increased to capture minute motion details and very fast speed like in sports [27]. IMUs can easily be used indoors, across long distances, and even underwater in some cases.

6.3.1 Vital signs

Vital signs can be defined as the measurements of the basic functions of the human body [187]. The four major vital signs that are always monitored by the medical doctors and healthcare specialists include [187]: The human body temperature, the pulse rate, the respiration/breathing rate, and the blood pressure. The blood pressure cannot be considered as a vital sign, however, it is often measured with the other vital signs. Vital signs are indeed useful in monitoring and detecting medical issues.

Epson sensing technology for vital signs monitoring is a product that has a sensing device that measures the vital signs indicating a biological activity in human bodies [188]. The proposed vital sign solution utilizes the photoelectric sensing. This measurement technology uses the hemoglobin's light absorbing properties. This technology is based on shining light using a green LED to the blood vessels lying under the skin. Epson has improved the measurement accuracy by using a 3-axial accelerometer and an algorithm to detect and eliminate the noise of the body movement (e.g., caused by arm movement).

In [189], the authors reviewed the utilization of wearable devices in healthcare and workout monitoring based on vital signals. This includes the electrocardiogram (ECG), electroencephalogram (EEG), electromyography (EMG), inertia, body movements, heart rate, blood, sweat, etc.

In [190], the authors reviewed the utilization of wearable devices and detection algorithms utilizing the piezoresistive/inertial sensors for monitoring the breathing factors and respiratory parameters based on inertial sensors (i.e., accelerometers and gyroscopes) and detecting dysfunctions or pathologies without surgery. Several tools were utilized to gather the respiratory parameters from the acceleration data.

In [191], the authors proposed an approach for detecting sleep parameters. This is achieved by measuring the earth magnetic field variation. The authors used a magnetometer sensor that is laid on the body which detects the night-time breathing movements (in millimeter). This is done by measuring the magnetic vectors changes. A wearable sensor that is soft and non-invasive is printed on a miniature printed circuit board. This includes a wireless Bluetooth low-energy (BLE) module and a low-power microcontroller. A minimum power consumption is achieved by utilizing a smart processing algorithm. This algorithm calculates respiration rate, apnea time, and time while movement.

In [192], the authors reviewed the conventional approaches for monitoring cardio-pulmonary rates and the ability of replacing these methods with radar-based approaches. The authors also studied the challenges that need to be overcame by the radar-based vital signs monitoring approaches in order to be applied in the healthcare domain. The authors presented a PoC for a radar-based vital sign detection and showed good measurement results. In [193], the authors highlighted the abilities of the IMU sensors to detect the vital signs. The results presented were promising in monitoring the soldiers' vital signs.

6.4 Ergonomics

Kuschan et al. [194] studied 2 lifting movements utilizing the acceleration and angular velocity signals. The ergonomic movement was done by the test subjects bending at hips and knees only generating the lifting power by squatting down. The unergonomic movement was done by the test subjects bending forward to lift the box with their backs mainly. The signals were collected by a vest equipped with 5 IMU sensors each with 9 dimensions. The IMU data collected from both the ergonomic and unergonomic movements were statistically analyzed and visualized. A CareJack vest was used that provided the worker with a real-time feedback about his ergonomic movement decreasing the spinal and back diseases risk.

Mudiyanselage et al. [195] proposed surface electromyogram (EMG) system using machine learning to recognize body movements that may harm muscles while handling materials. This study utilized a lifting equation developed by the U.S. National Institute for Occupational Safety and Health (NIOSH). The equation specifies a Recommended Weight Limit that recommends the maximum acceptable weight that a healthy worker can lift/carry and a Lifting Index value to assess the risk threshold. Four different ML models (Decision Trees, SVM, k-NN, RF) were developed to classify the risk assessments based on the NIOSH lifting equation. To find the best performance of each algorithm, the models sensitivity to different

parameters was also studied. The results showed that the Decision Tree model was able to predict the risk level with an accuracy near to 99.35%.

Humadi et al. [196] proposed a system based on IMUs and Kinect V2 for Rapid Upper Limb Assessment (RULA) risk assessment that was compared to a reference motion-capture cameras. The activities here were the different tasks of manual material handling. The calculated RULA scores were compared to the reference system using the proportion agreement index (Po) and Cohen's Kappa coefficient (κ). The wearable-based approach showed a "substantial" agreement ($\kappa > 0.6$) with the motion-capture cameras. The optical-based approach showed an "inconsistent" agreement with the motion-capture cameras ranging from "fair" ($\kappa < 0.4$) to "moderate" ($\kappa > 0.4$) agreements caused by self-occlusion and object-occlusion. Thus, the wearable-based approach was more suitable for in-field ergonomic risk assessment compared to the optical-based approach.

Fauziah et al. [197] proposed a tool to feedback with real-time assessment for workers ergonomic risks based on workers' postures. Direct visits to a manufacturing site in Indonesia were performed to study various production tasks to obtain complaints data of Work-Related Musculoskeletal Disorders (WMSDs). Tasks with a higher risk of injury were chosen to perform task analysis in order to segment the task steps. Appropriate risk assessment methods were chosen, e.g., Rapid Entire Body Assessment (REBA), Rapid Upper Limb Assessment (RULA), Ovako Working Posture Analysis System (OWAS), etc., based on the tasks critical characteristics. The tasks being studied included: cleaning up production residues, grinding processes, manual bending, and welding. The appropriate assessment method utilized was the OWAS. The measures considered as the WMSDs risk factors were the posture of back, arms, legs, and weight being handled. These measures were considered in the design of the sensor system which integrated 7 IMUs placed on the mentioned body locations. On a developed mobile application, a score shows the risk value and a warning signal is prompted to the worker.

Hsu and Lin [198] proposed a system of IMUs and exoskeleton robot for work performance assessment and assisting workers in performing different agricultural activities. By locating 3 IMUs on the worker's trunk and legs, the system controls the exoskeleton robot to move the workers' legs while supporting their backs. For evaluating the waist assistance system, ergonomic assessment tests were applied, e.g., Rapid Upper Limb Assessment (RULA) and electromyography (EMG). By using the exoskeleton robot, the muscle activity of the Erector Spinae was reduced by about 17.3% for lifting 10 kg objects during stoop lifting and by 20.4% for lifting 10 kg during semi-squat lifting. For the same tasks, the muscle activity of Biceps Femoris was reduced by 18.2% and 15.2%, respectively. This work is also applicable to rehabilitation activity, construction work, and other related work domains.

Humadi et al. [199] proposed an IMU system in order to assess the manual material handling activities using in-field RULA score. This was done using 3D Cardan angles and 2D projection angles relative to reference values recorded by a camera system for motion-capture. The axes conventions are generally used to establish the location/orientation of coordinate axes as a frame-of-reference. For the trunk and neck joint angles, the 2D convention had significantly smaller RMSE ($p < 0.05$). For the other upper-body angles, the convention had significantly smaller RMSE depending on the angle under analysis. The 3D convention showed a "moderate" agreement with the frame-of-reference, while the 2D convention showed a "substantial" agreement for two activities and a "moderate" agreement for one activity. For the 3D convention, the intra-class correlation coefficients ranged from 0.82 to 0.94. For the 2D convention, the coefficients ranged from 0.87 to 0.95 for repeated sessions performed by each subject. Thus, the proposed wearable IMU system using the 2D convention was found to be an accurate ergonomic risk assessment tool.

Jahanian et al. [200] proposed a system to study the risk and recovery metrics of arm use to discriminate between (1) Manual Wheel Chair (MWC) subjects having spinal cord injury (SCI) and matched able-bodied subjects and (2) MWC subjects with rotator cuff pathology progression for 1 year from users without pathology progression (longitudinal study). 34 MWC subjects and 34 age- and gender-matched able-bodied subjects were hired. Upper arm risk (humeral elevation $> 60^\circ$) and recovery (static ≥ 5 s and humeral elevation $< 40^\circ$) measurements were calculated from wireless IMUs placed on the upper arms and torso while being in a free-living environment. Two independent magnetic resonance imaging studies were evaluated for a subset of 16 MWC subjects about 1 year apart. The risk events frequency ($p = 0.019$), recovery events summed duration ($p = 0.025$), and each recovery event duration ($p = 0.003$) were higher for MWC subjects than able-bodied subjects. The risk events summed duration ($p = 0.047$), risk events frequency ($p = 0.027$), and risk to recovery ratio ($p = 0.02$) were higher and the recovery events summed duration ($p = 0.036$) and recovery events frequency ($p = 0.047$) were lower for MWC subjects having rotator cuff pathology progression ($n = 5$) with respect to the users without progression ($n = 11$). The IMU-derived measurements which quantified the arm use postures $> 60^\circ$ and risk to recovery ratios gave potential risk factors insights for rotator cuff pathology progression.

Vignais et al. [201] performed a material handling task ergonomic analysis by integrating a subtask video analysis and a RULA computation using a motion capture system combining IMUs and electrogoniometers. Five workers participated in the experiment and seven IMUs were located at the worker's upper body (pelvis, thorax, head, arms, forearms). An upper body biomechanical model continuously provided trunk, neck, shoulder, and elbow joint angles. Wrist joint angles were provided from electrogoniometers synchronized with the IMU system and the activity of the worker was video recorded at the same time. Joint angles were input to an ergonomic evaluation model during post-processing based on the RULA method. An RULA score was computed at each time step to specify the upper body exposure risk (right and left sides). Local risk scores were also calculated to identify the exposure anatomical origin. The video-recorded work activity was studied in time to recognize all subtasks included in the task. The mean RULA scores were for all subjects at high risk (for right and left sides, 6 and 6.2, respectively). A temporal analysis showed that workers spent at an RULA score of 7 most part of work time. Mean local scores showed that during the task most exposed joints were elbows, lower arms, wrists, and hands. Elbows and lower arms were during the work cycle total time at a high risk level (for right and left sides; 100%). Wrist and hands were exposed during most work period to a risky level (right: $82.13 \pm 7.46\%$; left: $77.85 \pm 12.46\%$). Subtasks called 'snow thrower', 'opening vacuum sealer', 'cleaning' and 'storing' were identified as the most unsuitable for right and left sides where the mean RULA scores and time spent percentages were at risky levels.

Villalobos and Mac Cawley [202] proposed an IMUs-based system to recognize human activity using AI and ML techniques to perform task classification and ergonomic assessments in workplace location. The authors used low-cost IMUs placed on slaughterhouse workers' wrists to gain information about their motions. The authors identified the wrists/hands risk factors leading to work-related musculoskeletal disorders (WRMSDs). The authors showed that by using low-cost IMU sensors on the slaughterhouse workers' wrists, the sharpness of the knife can be accurately classified and the worker RULA score can be predicted.

6.5 Human–computer interaction (HCI)

In [203], the authors proposed a wearable system that is operating in real-time for recognizing finger air-writing in 3-D space using the Arduino Nano 33 BLE Sense. This edge device runs TensorFlow Lite for action recognition on the device. The proposed system allows users to write characters (26 English lower-case letters and 10 digits) by moving fingers in air. This system uses a DL algorithm to recognize 36 characters from the motion data collected from the IMUs. This data is then processed by a microcontroller embedded in the Arduino Nano 33 BLE Sense. The authors prepared 63,000 stroke data samples written in air of 35 subjects (18 males and 17 females). This data was used in CNN training. The proposed system had a high recognition accuracy of 97.95%.

In [204], the authors proposed a hand gesture segmentation method that is a mobility-aware. The proposed system detects and segments hand gestures. The authors utilized a CNN in order to classify hand gestures considering noises due to mobility. The proposed system is called MobiGesture was used for healthcare. The authors used the leave-one-subject-out cross-validation test. The experiments with real subjects showed that the proposed approach achieves 94.0% precision, and 91.2% recall when the subject is moving. The proposed algorithm is 16.1%, 15.3%, and 14.4% more accurate than the state-of-the-art when the subject is standing, walking, and jogging, respectively.

In [205], the authors proposed two hand gesture recognition methods based on RNN. The first method is based on video signal and utilized a combined structure of CNN and RNN. The second method uses accelerometer data and only RNN. A fixed-point optimization is used to optimize the memory size for storing the weights and minimizing the power consumption in hardware and software processing. This optimization quantizes most of the weights into two bits.

In [206], the authors proposed FingerPing sensing technique. FingerPing is able to recognize different fine-grained hand poses. This is achieved by analyzing the acoustic resonance features. A surface-transducer that is fixed on a thumb ring is able to inject acoustic chirps (20–6000 Hz) into the body. There are four receivers on the wrist and thumb to gather the chirps. Various hand poses create different paths in order to the acoustic chirps to travel. This creates specific frequency responses at the four receivers. FingerPing could discriminate till 22 hand poses. This includes the thumb touching every phalange of the twelve one on the hand, also the ten American sign language poses. Experiments with 16 subjects had showed that the proposed system was able to recognize the aforementioned two poses sets achieving an accuracy of 93.77% and 95.64%, respectively.

In [207], the authors proposed a transfer learning approach to train target sensor systems from existing source sensor systems. This method utilizes system identification methods in order to learn a mapping function between the signals from the source sensor system to the target sensor system, and vice versa. This is doable for sensor signals that are of the same or cross modality. The

authors proposed two transfer learning models in order to translate recognition systems based on activity templates/models. This depends on the features of the source and target sensor systems. The proposed transfer learning approaches are evaluated in a HCI scenario; the transfer is done between wearable sensors located at various body locations, and between wearable sensors and ambient depth camera. The results showed that a nice transfer is achievable in just few seconds of data. This was done irrespective of the transfer direction for similar/cross sensor modalities.

In [208], the authors proposed a Quantum Water Strider Algorithm with Hybrid-Deep-Learning-based Activity Recognition that is called QWSA-HDLAR model. The proposed QWSA-HDLAR approach can recognize the various types of HAR. The QWSA-HDLAR model utilized a deep transfer learning approach. A neural-architectural-search-network (NASNet)-based feature extractor was used to generate feature vectors. The proposed QWSA-HDLAR method utilizes a QWSA-based hyperparameter tuning method in order to select the hyperparameters of the NASNet model. The classification of human activities is achieved using a hybrid CNN with a bidirectional RNN model. The experiments were done using two testing datasets, KTH and UCF Sports datasets. The results showed that QWSA-HDLAR model outperformed the state-of-the-art DL approaches.

7 Embedded HAR

In this section, we address HAR systems implemented on embedded devices. Basterretxea et al. [209] aimed at an FPGA-based autonomous single-chip HAR system without needing external computing means or human intervention. The system comprises all typical HAR process phases, i.e., signal segmentation, feature extraction through signal processing, feature selection through dimensionality reduction of input space, and activity recognition using a neural network classifier. A physical activity recognition application was utilized as a frame-of-reference to evaluate the system performance and analyze FPGAs potential benefits in prospective wearable HAR applications.

Czabke et al. [210] presented a microcontroller-based algorithm for human activity classification based on a 3-axial accelerometer. For long term activity patterns monitoring, the data amount is kept small with efficient data processing. The authors classified in real-time the following activities: resting, walking, running, and unknown activity. The algorithm ran on the proposed device Motionlogger with a key fob size and can be placed in a pocket or a handbag. The algorithm testing with 10 users locating the Motionlogger in their pockets showed an average accuracy $> 90\%$.

Li et al. [211] presented an activity recognition system from passive RFID data using a deep convolutional neural network. For activity recognition, the RFID data are feed directly into a deep convolutional neural network, instead of selecting features. A cascade structure is used that first detects from the RFID data the "object use" then predicting the activity. The "object use" of specific activities means recognizing activities based on the used objects, i.e., detecting human–object interaction from the sensed data. The proposed system solves activity recognition as a multi-class classification problem, thus it is scalable for applications having big number of activity classes. Testing the proposed system using RFID data obtained in a trauma room comprising 14 h of RFID data from 16 actual trauma resuscitation (restoration of acute illness or near death). The proposed system outperformed state-of-the-art systems implemented for activity recognition and had similar performance with process-phase detection like systems based on wearable sensors. The authors also studied the strengths and challenges of the proposed deep learning architecture for activity recognition using the RFID data.

Sundaramoorthy et al. [212] proposed HARNet that is efficiently handling resources and computationally practical network for on-line Incremental Learning and User Adaptability to be a mitigation technique used for anomalous user behavior recognition. Heterogeneity Activity Recognition Dataset [152] was utilized for evaluating HARNet and other proposed variants by using acceleration data collected from different mobile platforms. The authors utilized Decimation for down-sampling in order to generalize sampling frequencies between mobile devices. Discrete Wavelet Transform was utilized for preserving information among frequency and time. Typical evaluation of HARNet on the adaptability of users resulted in an accuracy increase by 35% by utilizing the model ability to extract differentiating features between activities inside heterogeneous environments.

Saponas et al. [213] argued that researchers have explored various approaches like services running in mobile phone background, cameras in environment, RFID readers, etc. The authors recommended that for HAR system to transition from interesting research topic to mainstream technology and for enhancing the day-to-day activities, HAR must use many of the user's computing ecosystem components, e.g., low-power embedded hardware running continuously and mounted on phone, cloud services tracking the user's activities on long-term and recognizing current activity from all inputs, etc. The authors provided a brief discussion of how this system can be implemented and utilized.

Chen et al. [214] proposed a multi-level HAR modeling workflow named Stage-Logits-Memory Distillation (SMLDist) in order to construct deep convolutional HAR

models using embedded hardware. SMLDist comprises stage distillation, memory distillation, and logits distillation. Stage distillation limits the intermediate features learning direction. The student models were taught by the teacher model how to explain and save the inner relationship between high-dimensional features depending on Hopfield networks in memory distillation. Logits distillation constructs logits filtered by a smoothed conditional rule to keep the probability distribution and improve the accuracy of the softer target. The authors made a comparison of the accuracy, F1 macro score, and energy cost on a MobileNet V3 model [215] embedded platforms built by SMLDist with different state-of-the-art HAR systems. The batch size of the training and validating samples was set to 256 while the learning rate was set to 1×10^{-4}. The proposed model had a good balance to be robust and efficient. SMLDist can as well compress the models given a slight performance loss with an equal compression ratio compared to other advanced knowledge distillation approaches on 7 public datasets.

Bhat et al. [216] proposed the first HAR system performing both tasks of online training and inference. The proposed system first generates features using the FFT and discrete wavelet transform of a stretch sensor that is textile-based and accelerometer. Based on these features, the authors developed an ANN classifier that is trained online utilizing the policy gradient algorithm. The experiments on an IoT device (TI-CC2650 MCU) that is low power with nine subjects showed an accuracy of 97.7% in recognizing 6 activities and the transitions between these activities with < 12.5 mW consumption of power. Another line of work by Bhat et al. [217] introduced an open source platform called OpenHealth that allows for health monitoring using wearable sensors. OpenHealth represents a hardware/software set and wearable devices enabling autonomous integration of clinically related data. The OpenHealth platform comprises a wearable device, standard software interfaces and implementations of applications for activity and gesture recognition.

Alessandrini et al. [218] solved the HAR problem directly on embedded devices by developing an RNN on a low cost, low power microcontroller, while checking the targeted performance through accuracy and low complexity. The authors first implemented an RNN which combines Photoplethysmography (PPG) and 3-axial accelerometer data in which these data is used to match motion artifacts in PPG to recognize human activity accurately. The authors then deployed the RNN to a Cloud-JAM L4 embedded device that is based on an STM32 microcontroller. This device is optimized to preserve an accuracy $> 95\%$ whilst requiring small computational power and memory resources. The experiments showed that this system can efficiently be developed on a limited-resource system, aiming at designing a wearable embedded system that is fully autonomous for HAR and logging.

Choudhury et al. [219] presented a Mobile Sensing Platform (MSP) that is a small wearable device. MSP is designed for accurate activity recognition on embedded devices to support ubiquitous computing applications that are context-aware. The authors had performed many real-world deployments and subject studies. The results are utilized to enhance the hardware/software design and HAR algorithms.

Ravi et al. [220] a HAR system based on a deep learning for accurate and real-time classification on wearable devices that have low power. In order to gain steadiness against changes in the orientation, placement, and sampling rates of sensor, a feature generation using the IMU data spectral domain. The proposed method utilized the transformed input temporal convolutions sums. the proposed approach accuracy is assessed versus the current state-of-the-art techniques on both laboratory and real world activity datasets. The authors performed a systematic analysis for the parameters used in feature generation. In addition, the authors compared the activity recognition computation times on mobile devices and sensor nodes. The proposed system was deployed as an Android app and as an embedded algorithm as well for the Intel Edison Development Platform. Intel Edison was used to show human activity classification on-node by using the trained recognition model. Using a dual-core Intel Atom CPU with rate of 500 MHz, wireless connectivity, and compact physical with dimensions of $35.5 \times 25.0 \times 3.9$ mm, the Intel Edison is a small powerful platform that is suitable for smart sensor networks applications.

Table 14 presents a comparison and summarize several embedded HAR systems, including power consumption, performance metrics and time consumption.

8 Discussion and conclusion

Human activity recognition (HAR) has gained tremendous momentum in recent years along with the digitization of all aspects of society in general and individual lives in particular. Of particular significance and potential are those systems based on streaming of inertial data (such as linear and rotational motion) from IMU devices. Such sensing capabilities are currently embedded in almost all commodity mobile and wearable devices. As the field is very huge and diverse, this article is by no means comprehensive; it is though meant to provide a logically and conceptually rather complete picture to advanced readers, as well as to present a readable guided introduction to newcomers. Our logical and conceptual perspectives mimic the

Table 14 Comparison of several embedded HAR systems, including power consumption, performance metrics and time consumption

Reference	Power consumption	Performance metrics	Time consumption
Basterretxea et al. [209]	268 mW (system running at maximum speed)	Overall accuracy: 84.9% (Minimal ANN structure) and 89.7% (Augmented structure)	2.5 s
Czabke et al. [210]	–	Average accuracy $> 90\%$	5.7 s
Li et al. [211]	–	Average accuracy 72.03% for 5 resuscitation phases detection, 80.40% for 11 activities recognition, and 90.8% for 6 lab activities recognition	–
Sundaramoorthy et al. [212]	–	Accuracy of HAR-LCNet (LSTM $\rightarrow$ Conv-1D $\rightarrow$ Conv-2D) is 96.79%	Time taken for Classification of a single window of HAR-LCNet is 68.9 ms
Chen et al. [214]	Energy cost of SMLDist (Stage-Memory-Logits Distillation) ranges from 0.0209 to 0.0385 (mW $\cdot$ h/sample) on 7 public datasets	Accuracy of SMLDist ranges from 92.11 to 99.41% on 7 public datasets	Time cost of SMLDist ranges from 8.653 to 14.626 (ms/sample) on 7 public datasets
Bhat et al. [216]	Computation power consumption is 11.24 mW. Sensing power consumption is 1.13 mW. BLE Communication power consumption is 5 mW.	Accuracy of 97.7% in identifying 6 activities and their transitions	Processing time is 27.60 ms. Sensing time is 1500 ms. Bluetooth Low Energy (BLE) communication time is 8.6 ms.
Bhat et al. [217]	Power consumption is about 12.5 mW for single HAR activity recognition. Power consumption is about 10 mW for gestures recognition.	Accuracy is $> 90\%$ for all HAR activities. Accuracy is 98.6% for gestures recognition.	–
Alessandrini et al. [218]	–	Accuracy is over 95%	Average test time is 150.1 ms at the highest test accuracy of 95.54% on the micro-controller unit (MCU)
Choudhury et al. [219]	Power consumption is about 43 mW (with all sensors running continuously)	Highest Accuracy is 93.8% (Inference v2.0 with temporal information included and supervised learning)	Activities classified every 0.25 s
Ravi et al. [220]	–	Accuracy is 95.1% on ActiveMiles dataset, 98.2% on WISDM v1.1, 95.9% on Skoda (Node 16), about 95% on Daphnet FoG	Computation time is 5.4 ms/20 ms/13–42 ms for Spectrogram generation and 5.7 ms/8 ms/14.9 ms for Deep Learning classification on LG Nexus 5/Samsung Galaxy S5/Intel Edison, respectively.

typical data science pipeline for state-of-the-art AI-based systems. Namely, starting with datasets, either publicly available and/or collected, to feature engineering and/or automatically extracted, developing learning models including the transfer of pre-trained models to new datasets and/or tasks, and finally validating the trained models. We have as well touched on adversarial attacks as actual deployment involves security/privacy issues. Adversarial attacks against timeseries data is a pretty recent research and only a handful of works are available.

There are huge opportunities both on the research and industrial sides for IMU-based HAR. To the best of our knowledge, there have been very few works addressing the fusion of inertial and other data modality such as the visual. Some recent work in empirical psychology aim at merging IMU with social media texts trying to profile and understand the mental status of the user and potential risks with behavioral patterns and mood changes, for example. With the COVID-19 outbreak, gyms among other public places have been forced to close their doors in order to control the spread of the global pandemic. Moreover, people were advised to stay at home and in some places, were fined if found outside. Hence, many people's mental and physical state started to deteriorate. As a result, more and more workout mobile apps started to gain an increasing popularity. Accordingly, there has been a growing interest in

performing exercising and sports indoors with the assistance of smart digital technology for tracking and assessment of such exercising. Such kinds of apps are based essentially on measuring and tracking motions, and here where inertial data are readily, naturally, and directly helpful. Much more quality inertial datasets need to be collected and made open, where the setup and data configuration have to be well documented. We believe that for closed HAR systems (deployment environment very much resembles the training model construction environment), deep neural methods may not be that effective, especially with the limitation of datasets and their sizes. However, as seen above, deep methods are very effective and useful in transfer learning handling the deployment of HAR systems in radically different situations and contexts from those where the model was built.

Inertial data streaming through IMU devices are more intrinsic than other data modalities, specially the visual, as they are embedded into wearables and hence, mobile carried on with the user wherever she is. It has as well the advantage of complying more with the security and privacy concerns of the users. Its main disadvantage is high sensitivity to noise and low explicit semantic content, as compared, for example, with visual, audio, or textual data. Hence, they need smart algorithms to extract the most feasible and valuable information from them.

Some research works have applied advanced 3D data tensor models in medical images analysis, specifically the Azheimer's disease [221–224]. However, the modalities used in these works are visual scans (such as magnetic resonance and in general 3D visual data). To the best of our knowledge, these models have not been used for analysis of inertial motion data which is the main focus of the current paper. However, it is worth investigating the use of such architectures, and others like normalizing flows and GANs (Generative Adversarial Networks), for the analysis of timeseries inertial motion data.

Another crucial research point in the domain of HAR based on IMU Data is the "adversarial attacks" on time series data. Deep neural networks are highly parameterized complex models and they can be highly sensitive, and hence, ill-robust, to small perturbations, as observed by [225]. This is most witnessed in visual data, where the perturbations can be completely imperceptible to the human eye, however, may lead to vastly and widely different classification outputs than expected or desired [225]. These perturbations may be explicitly engineered (target adversarial attack) and added to inputs for *malicious* purposes, which can potentially have devastating consequences in critical production systems. Therefore, there is a need to design methods that are robust to such attacks, detecting and mitigating them, in order to maintain the integrity and correct functionality of such systems [226].

Most of the work done in that direction have targeted visual data, aiming at the detection and mitigation against such attacks including techniques such as *adversarial learning* and the proper development of adversarial examples. Attacking timeseries streaming data can be of detrimental consequences as well. For example, models that are deployed to detect and monitor seismic waves can be fooled into creating fear and panic in society; wearables used for activity recognition from inertial signals can fool the user into believing performing fake actions, or believing some performance level in a given activity [227].

Only recently there have been some emergent works targeting adversarial attacks against timeseries data, among which are the data streamed from IMU sensors. To the best of our knowledge, Fawaz et al. [228] in 2019 were the first to address adversarial attacks against deep learning models trained on timeseries data. Such timeseries adversarial examples can severely affect the predictive performance and reliability of decisions taken by such models in real-world applications specially those related to healthcare applications, sports, and security. The crux of this work is that timeseries data, similar to visual data, are also prone to adversarial attacks. And even they developed adversarial attacks (examples) by transferring and adapting adversarial attacks that had been shown to work on images. More specifically, the authors formulated attacks on timeseries data based on two methods: (1) the Fast Gradient Sign Method (FGSM) and (2) the Basic Iterative Method (BIM). FGSM is based on one step update along the direction of the gradient at each time step (this is the opposite to the learning update).

Saponas et al. [213] highlighted the key challenges which are facing the adoption of HAR in future systems in industry. The authors mentioned that performing HAR using "mobile devices" is the crucial factor and accordingly the authors highlighted two important challenges: the location independence and continuous operation/function. The authors mentioned that no single device can figure all the context needed for the required HAR use cases. The

authors argued that the future HAR systems have a challenge to utilize all the computing ecosystem components from embedded devices to the cloud.

In [229], the authors argue that one might have the false perception that performing HAR using IMUs is already at a perfect state and that no further actions need to be done. On contrary, out of those who are not using their wearable devices sensing: 36% cited that this is due to the measurement accuracy and 34% cited that is because the improper activity tracking [230]. The authors argue that HAR is still far away from being solved due to the following reasons: (1) the difficulty in defining an activity, (2) Heterogeneous benchmarks, (3) Collecting ground-truth data for the HAR systems is expensive and not easy, (4) Various characteristics causes diverse activity profiles, (5) the lack of standardization in storing, processing, and analyzing the data. These points highlight the open research problems in this field and give some insights about the future research works.

Appendix A. Public activity datasets

There is a wide variety of activity datasets that are based on inertial motion data. These are typically used as benchmarks to test new techniques for activity analysis and inference systems. A brief description of these datasets are given in Table 15.

Table 15 Brief description of public activity datasets (F refers to female and M to male)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Food preparation activities					
Pham and Olivier [231]; slice and dice	20	11 actions involved in food preparation of a salad and sandwich; chopping, peeling, slicing, dicing, coring, spreading, eating, stirring, scooping, scraping, and shaving	3 knives and a serving spoon with embedded modified Wii Remotes	A 3-axis ADXL330 accelerometer is integrated in the Wii Remote	A vector of 80 features was used for training the classifiers; overall accuracy of 82.9%
De la Torre et al. [232]; CMU-MMAC	32	Food preparation, Cook five recipes: Pizza, Sandwich, Brownies, Scrambled eggs, Salad	Left and right forearm, arms, calves, thighs, wrists, abdomen, and forehead	Camera, microphone, RFID, 3D gyroscopes, 3D accelerometers, 3D magnetometer, ambient light, heat flux sensor, galvanic skin response, temperature, motion capture	The dataset focuses on cooking activities and it is obtrusive due to many sensors on human body
Gait and walking activities					
Huang et al. [233]; The diverse gait dataset	22 (9 F, 13 M)	4690 walking strides data and 19, 083 gait labels	Each user wore an Xsens MTw IMU (3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer). The measurement unit was placed at the in-step on the shoe right side	Inertial motion data recorded by IMU and foot motion images recorded by phone camera	Age range 32.5 ± 7.5 years. threshold-independent stride segmentation algorithm SDATW; F-measure of 0.835
Makela et al. [234]; VTT-ConIoT	13 (3 F, 10 M)	16 activities; painting, spraying pain, spreading screed, vacuum cleaning, picking up trash, climbing up stairs, working with hands up, jumping down from stairs, working on high tubes, working on low tubes, working on floor crouching, working on floor kneeling, pushing cart with objects, walking on straight line, walking around objects, stairs up-down	1 sensor at hip, 2 sensors near non-dominant hand shoulder; on arm upper part, on shoulder back part	Aistin iProxoxi sensors; 10-DOF IMU; 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer, barometer	Age range 25–55 years. classification accuracy of up to 89% for six class setting, up to 78% for more granular sixteen class setting
Alemayoh et al. [235]	8 (0 F, 8 M)	8 activities; walking, jogging, jumping, ascending stairs, descending stairs, still (standing and sitting), lying down, bicycling	Waist	iPhone 7 smartphone; 6 IMU sensors (three 3-axial accelerometers and three gyroscopes)	About 150 min of recorded data. Double-channel Convolution Neural Network (DC-CNN); overall accuracy 99.40%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Zhou et al. [236]; Personal-specific gait recognition	9	5 gait patterns; walking, upstairs, downstairs, slope up, slope down	6 IMUs; left/right thighs, left/right knees, left/right feet	3-axial accelerometer and 3-axial gyroscope	LSTM-based human gait recognition model and CNN-based personal-specific disentangle method; average accuracy 87%
Kim et al. [237]; Fine-grained walking patterns	36	18 walking styles; walking on flat ground, walking on snow, walking with umbrella in left/right hand, walking while looking at smartphone in left/right hand, walking with dumbbell in left/right/each hand, walking with arms crossed/hands behind back/hands in pockets, jogging, uphill, downhill, upstairs, downstairs, standing and sitting	Smartwatch worn on the non-dominant hand (i.e., left wrist)	Sensing device: smartwatch (DW9F1 by Fossil Group); 3-axial accelerometer and gyroscope. Host device: smartphone (Galaxy Note 20 by Samsung Electronics)	Age range 20–62 years, LSTM overall accuracy 97.158% LSTM with attention overall accuracy 97.096%
Li et al. [238]; Alcogait 2.0: blood alcohol concentration from smartphone gait data	Over 100 gender ratio (F:M) 8:5	Detecting intoxication levels of alcohol drinkers from their gait	Smartphone kept in the left back pocket. Watch placed on the left wrist	A Google Pixel XL and LG Watch Sport: accelerometer and gyroscope	Age range 21–65 years, weight range 38.56–104.33 kg, Bi-LSTM and CNN predict blood alcohol concentration (BAC); Bi-LSTM RMSE is 0.0167 and CNN RMSE is 0.0168
Madcor et al. [239]; VSGD	45 (13 F, 32 M)	Gait: walking a straight path while wearing 4 IMUs and being filmed using 2 smartphones fixed at 2 different directions	3 MetaMotionR (MMR) devices on the right upper arm, the right thigh, and the right knee, whereas the Apple watch worn on the right wrist	4 IMUs; 3 MMR and an Apple Watch Series 1. 2 iPhone 7 cameras with Quality 1080P and with 30 frames per second	Ages range 18–23 years
Fuller et al. [240]	49 (26 F, 23 M)	6 movement types: lying, sitting, self-paced treadmill walking, treadmill walking/running at 3 Metabolic Equivalents of Task (METs), 5 METs, and 7 METs	Fitbit placed on one wrist and Apple Watch placed on the opposite wrist	Apple Watch and Fitbit; heart rate, steps, distance, and calories	Subjects recorded 65 min; 40 min treadmill, 25 min sitting/lying. Tenfold cross validation. Rotation Forest has highest accuracy for Apple Watch 82.6%, Random Forest has highest accuracy for Fitbit at 90.8%
Permatasari et al. [23]; MMUISD gait database	299 (53 F and 246 M)	Walking on flat ground corridor; 3 different speeds; slow, normal, fast	In hand (left/right), trouser pocket (left/right), back-pack, handbag	3-axial accelerometer, 3-axial gyroscope	Age range 18–28 years, Each subject in 5–8 min performs all walking speeds; slow, normal, fast. Tenfold cross validation. Logistic Regression, LDA, MLP, Decision Tree, RF, Gaussian Naive Bayesian, Gradient Boosting, SVM; RF achieves the highest accuracy 53.87% (walking speed "normal" + all sensor positions)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Fang et al. [241]	9 (4 F, 5 M)	5 gait patterns; walking, upstairs, downstairs, slope up, slope down	7 IMUs; waist, left/right thighs, left/right knees, left/right feet	3-axial accelerometer and 3-axial gyroscope	Gait Neural Network (GNN) for Human–Exoskeleton Interaction; based on temporal convolutional network; gait recognition overall accuracy 98.81%
Mostafa et al. [19] and Adel et al. [20]; EJUST-GINR-1	20 (10 M, 10 F)	Gait	IMUs mounted on different 8 body locations, two of them are embedded sensors in Apple smartwatches worn on the right and left wrists, see Fig. 9	3D accelerometers and 3D Gyroscopes	Equal number of males/females (each 10), ages range 19–33, heights range 146–187 cm, and weights range 56–130 kg
Zhang et al. [242]	40	5 indoor activities; walking, stillness, stair climbing, escalator, elevator taking	4 smartphone placement; trouser pocket, bag, hand, holding horizontally	Accelerometer and gyroscope for HAR, and barometer to better recognize the altitude changes for different activities	The average recognition F-score of XGBoost reaches 84.41%
He et al. [243]; Weakly labeled dataset	10	5 ADLs; walking, jogging, jumping, going upstairs, going downstairs	Right trouser pocket	iPhone; 3-axial accelerometer	91, 266 sequences; each 40.96 s. supervised learning and reinforcement learning agent interacts with sensor data over time. accuracy; CNN baseline model 85.17%, DeepConvLSTM 92.31%, the proposed model 95.35%
Malekzadeh et al. [244]; Motion sense	24 (10 F, 14 M)	6 activities in 15 trials in the same environment: downstairs, upstairs, walking, jogging, sitting, and standing	Smartphone located in the front pocket	iPhone 6 s: accelerometer (acceleration and gravity), attitude (pitch, roll, yaw), gyroscope	A total of 24 participants in a range of gender, age, weight, and height
Genovese et al. [245]	16 (8 F and 8 M)	Step counter; slow and intermittent ambulation	Waist-worn	3-axial accelerometer; Gear 2 smartwatch; custom wearable app ADAM	Training dataset; 8 subjects (3 F, 5 M); age range 28–55 years, height range 160–185. testing dataset; 8 subjects (5 F, 3 M); age range 29–54 years, height range 158–187. normal walking over long paths; mean absolute relative error 1–3%. Slow and intermittent ambulation; error is limited to 5%
Khandelwal et al. [246]; MAREA	20 healthy subjects (8 F, 12 M)	Gait; walking and running in indoor and outdoor environments	Waist, left wrist, both ankles	3-axes Shimmer3 accelerometers	Age range 33.4 ± 7 years, mass range 73.2 ± 10.9 kg, height range 172.6 ± 9.5 cm. Controlled indoor environment; combined median F1-Score: 0.98 for Heel-Strikes, 0.94 for Toe-Offs. Outdoor street; combined median F1-Score: 0.82 for Heel-Strikes, 0.53 for Toe-Offs
Esfahani et al. [247]; PAMS	20	Sitting, standing, walking, running, ascending/descending stairs, cycling	Smartphone in hand or in pants pocket (the same side or the opposite side)	Smartphones; e.g., Galaxy S7, Galaxy S6, Galaxy S4, Sony F5122, and Xiaomi Redmi. Sensors: accelerometer and gyroscope	The average precision for all activities is above 88.5%. age range 21–89 years

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Chereshnev et al. [22]; HuGaDB	18 (4 F, 14 M) healthy subjects	Walking, running, stairs up and down, sitting down, etc	IMUs sensors placed on the right and left thighs, shins, and feet. Electromyography sensors placed on the quadriceps (front thigh)	6 wearable MPU9250 inertial sensors (3-axis accelerometer and 3−axis gyroscope) and 2 electromyography (EMG) sensors to measure muscle activity	10 h of recording; age average 23.67, std 3.69 years; height average 179.06, std 9.85 cm; weight average 73.44, std 16.67 kg
Dutta et al. [248]	122	8 activities; "Tall male 400 m walk on hard surface", "Tall female 400 m walk on hard surface", "Tall male 400 meters run on hard surface", "Tall female 400 m run on hard surface", "No disability 0.25 miles hard surface", "Short Physical Performance Battery 0.25 miles hard surface", "Cane 100 m hard surface", "Walker 100 m hard surface"	Dominant wrist	3-axial accelerometer	99 seniors (age above 65 years; 24 healthy, 75 disability). 23 tall participants (males; average height 188.55 cm, average weight 102.41 kg, females; average height 180.39 cm, average weight 79.3 kg) GMM; classify 8 activities; overall accuracy 88.92%. HMM; classify 7 activities; overall accuracy 93.5%
Stisen et al. [152]	9	6 activities: biking, sitting, standing, walking, stair up, and stair down	Each subject is carrying 8 smartphones kept in a tight pocket carried around the waist, and 2 smartwatches worn on each arm	Accelerometer/gyroscope sensors across 36 Android and iOS smartphones and Android smartwatches (31 smartphones, 4 smartwatches, 1 tablet) spanning over 13 different models from 4 manufacturers	Age range 25–30 years. Mobile Sensing Heterogeneities for Activity Recognition. Each subject is carrying 8 smartphones (2 instances of 4 smartphone models) and 4 smartwatches (2 instances of 2 smartwatch models)
Bruno et al. [249] ; Human odometry outdoor dataset (HOOD)	1	6 motions types; slow walking, normal walking, running, slow crawling, fast crawling, slithering. 6 outdoor environments; grass field, uphill road, staircase, river bed, woods, snow. 2 different path configurations: straight line, zig-zag	Right foot, waist, chest, right wrist	InvenSense MPU6050 IMU: 3-axial accelerometer, 3-axial gyroscope	
Gupta et al. [250]	20	Walking, running, sitting, standing	Jeans pocket	ASUS ZenFone 5 Android smartphone; 3-axial accelerometer	SVM, J48, AdaBoost, Random Forest; AdaBoost achieves best accuracy 98.8283%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Ngo et al. [43]; OU-ISIR	744	Level walking, up-slope walking, down-slope walking	Waist (belt mounted, 1 IMU on each side of waist, 1 IMU and a smartphone at center back waist)	3 IMUs and one smartphone (Motorola ME860); 3-axial accelerometer and gyroscope of IMUs and 3-axial accelerometer of smartphone	Age range 2–78
Bayat et al. [251]	4 (2 F, 2 M)	6 activities; running, slow walking, fast walking, aerobic dancing, stairs-up, stairs-down	Phone in hand, phone in pocket	Android phone: 3-axial accelerometer	Age range 29–33 years. average of probabilities of three classifiers (Multilayer Perception, Logit Boost, SVM); overall accuracy 91.15% when phone in hand
Reyes et al. [42, 91]; UCI HAPT	30	6 basic activities: 3 static postures (standing, sitting, lying) and 3 ambulation activities (walking, walking downstairs, walking upstairs). 6 postural transitions; stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand	Samsung Galaxy S II Smartphone placed on the waist	3D accelerometer and 3D gyroscope	Age range 19–48 years. It is called 'The Human Activity and Postural Transitions' (HAPT) and collected using Android smartphones at a sampling rate of 50 Hz
Shoaib et al. [67]	10 (10 M)	7 activities; walking, running, sitting, standing, jogging, biking, walking upstairs, walking downstairs	Right/left jeans pocket, belt position towards right leg, right upper arm, right wrist	5 smartphones (Samsung Galaxy SII i9100); accelerometer (with and without gravity), gyroscope, magnetometer	Each activity is performed for 3–4 min. Age range 25–30 years. Bayesian networks, SVM, Logistic regression, k-NN, Decision trees, Random forest. 3 feature sets consist of time domain features, 1 consists of frequency domain features
Aung et al. [252]	8	2 gait events; heel strike, toe off	Heel, ankle, shank, waist	Camera-based motion tracking Vicon system, 2 3-axial accelerometers	3 datasets; Set A: simulated accelerometer data for comparisons between various sensor locations. Sets B and C: actual accelerometers for experimentation of various types of walking terrain. Wavelet transform + feature reduction using manifold embedding + Gaussian mixture model (GMM); Leave-one-out (LOO) cross-validation at subject level: Set A: average F1 scores range 75–86%. Sets B and C: average F1 scores over 90%
Do et al. [133]	8	4 activities; walking, staying still (the phone is stationary), running, driving	No restriction on the way of carrying the phone (e.g., in hand bag)	Smart phone 3-axial accelerometer and GPS	Age range 6–67, ANN trained on a set of 6850 accelerometer samples
Casale et al. [253]	20 (5 F, 15 M)	7 walking scenarios; indoor corridor, outdoor street uphill/downhill, crowded/free flat urban street, crowded pass through doors from urban to indoor environment, semi-indoor corridor with up/down ramps, walking in garden with rough floor	Mobile phone placed in jacket pocket on chest	Google Nexus One Android mobile phone; 3-axial accelerometer	Total of 140 different walking runs. age range 25–35 years. AdaBoost classifier and tenFold cross-validation; False Rejection Rate (FRR) = 0.003 and False Acceptance Rate (FAR) = 0.0001

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Zhang et al. [12]; USC-HAD	14 (7 F, 7 M)	12 activities; walk forward, walk left, walk right, walk up-stairs, walk down-stairs, run forward, jump, sit on chair, stand, sleep, elevator up, elevator down	Front right hip	3D accelerometer and 3D gyroscope	Data taken from one sensor location but it focused on ambulation activities. See Fig. 3a for data collection setup
Lara and Labrador [254]	10 (2 F, 8 M)	Running, walking, sitting	Chest sensor strap	HTC Evo 4 G mobile phone; accelerometer, GPS, BioHarness^{TM}BT chest sensor strap; heart rate, respiration rate, breath amplitude, skin temperature, posture/ inclination, electrocardiogram amplitude, galvanic skin response, SpO_2, 3-axial accelerometer	2 new subjects (1 F, 1 M) that were not in training phase 8 (1 F, 7 M) [255] performed each activity sequentially for 5 min. C4.5 Decision tree; overall accuracy 92.64%
Czabke et al. [210]	10 (3 F, 7 M)	Activities; walking, running (all the subjects excluding one with high heels sandals) on various ground (outside), resting (sitting or standing). undefined activity: moving but not walking or running	Trousers Pocket	3-axial accelerometer	Age range 55–22 years; age average 25 years. Total more than 50 min activity data recorded. Algorithm runs on ATmega 644 microcontroller. Threshold-based, overall accuracy 90.0%
Bovi et al. [256]	40 healthy subjects (22 F, 18 M)	Multiple-task gait analysis protocol: walking at different speeds, walking on toes, walking on heels, step ascending, step descending	29 retro-reflective markers located on head, upper limbs, trunk, pelvis, lower limbs	9-cameras SMART-E to capture motion. 2 force plates for ground reaction forces (GRFs). Electromyography signals (EMG) with 8-channel wireless electromyograph	A multiple-task gait analysis approach. adult group: 20 aged 22–72 years, mean 43.1 ± 15.4, body mass 68.5 ± 15.8 kg, height 1.71 ± 0.10 m, 11 F, 9 M. young group: 20 aged 6–17 years, mean 10.8 ± 3.2, body mass 41.4 ± 15.5 kg, height 1.47 ± 0.20 m, 11 F, 9 M
Kwapisz et al. [106]; WISDM	29	6 daily activities; walking, jogging, ascending stairs, descending stairs, sitting, standing	In the front trouser pocket	Accelerometer in several types of Android smartphones; Nexus One, HTC Hero, and Motorola Back-flip	Multilayer Perceptron Neural Network; overall accuracy 91.7%
Longstaff et al. [257]	22	Staying in one place, walking, running	Smartphones held in hand or worn on hip (belt or pocket) or armband	HTC Android Dev Phone 1; GPS receiver, Accelerometer	30 min of each activity; total 90 min per subject. Tenfold cross-validation. NB, DT; Use of co-learning
Santos et al. [258], UPCASE	1 (1 M)	Walking, running, idle, resting	Mounted on backpack/on vest	Sony Ericsson W910i smartphone/Nokia N95 smartphone, and BlueSentry sensor; two 3-axial accelerometers	500 sensor readings for each activity in 4 different environments. Decision trees; overall accuracy; ID3 (96.13%), C4.5 (95.82%)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Krishnan et al. [259]		Lower body activities (walking, sitting, standing, running, lying down)	Right ankle and left thigh	Two 3-axial accelerometers	AdaBoost algorithm built on decision stumps; overall accuracy 95%
Wang et al. [260]	52 (13 F, 39 M)	5 walking patterns; walking flat, walking slope-up, walking slope-down, ascending stairs, descending stairs	Mounted at waist above iliac spine	3-axial accelerometer	Age range 21–64 years, height range 1.53–1.88 m, weight range 42–94 kg. Each walking pattern performed 10 times each for 11–29 s. Using random frame selecting (RFS) train-test method. Multilayer Perceptron Neural Network; overall accuracy 92.05%
Coley et al. [261], First study	10 (3 F, 7 M)	Stair ascent, stair descent, flat walking	Gyroscope attached to shank, ultrasonic system at ankle lateral malleolus and knee lateral femoral	Gyroscope, ultrasonic motion system; 3 fixed loudspeakers (emitters), 2 microphone markers (receivers)	Age range 24 ± 4 years. Each on-body marker 3D spatial positions are estimated by triangulation. Shank anterior–posterior rotation and derivative are estimated by microphone markers spatial coordinates. Overall sensitivity and specificity of stair ascent; 98.2% and 97.5%
Coley et al. [261], Second study	3	Flat walking 25 m, 50 stairs ascent, stair descent	Attached to shank	Gyroscope	Elderly subjects; age range 81 ± 3 years. Number of steps range 11–22. Overall sensitivity and specificity of stair ascent; 97% and 99.3%
Coley et al. [261], Third study	7	196 walking upstairs cycles	Attached to shank	Gyroscope	Elderly subjects; age range 61 ± 8 years. Record 60 min during daily physical activity. Overall sensitivity and specificity of stair ascent; 100% and 100%
Lee and Mase [262]	8 (2 F, 6 M)	5 activities; sitting, standing, walking on level ground (slow, normal, fast), walking upstairs, walking downstairs	Right or left trouser pocket	2-axial accelerometer and gyroscope	Age range 23–51 years. Threshold-based and Fuzzy-logic reasoning method. recognition ratios; walking on level ground 95.91%, walking upstairs 94.35%, walking downstairs 92.85%
Randell and Muller [263]	10	6 activities; walking, running, sitting, walking upstairs, downstairs, standing	Trouser pocket	One 2-axial ADXL202 accelerometer	Neural Network; overall accuracy; 85–90%
Sports, Fitness, Exercises					
Sharshar et al. [27]; MM-DOS	50 (50 M)	4 workout exercises: free squats, shoulder press, push ups, lunges	IMUs; left upper arm, right upper arm, left wrist, right wrist, left calf, right calf, left thigh, right thigh, lower back	Kinect v2; color camera, depth sensor (infrared camera/infrared projector), microphone array. 9 IMU units; from each IMU, authors used 3-axial accelerometer, 3-axial magnetometer, 3-axial gyroscope, pressure, temperature. 2 RGB cameras. Thermal camera Flir One Pro	Age range 17–30 years, height range 160–195 cm, weight range 60–130 kg. CNN/SVM; accuracy up to 96% for RGB videos, 90% for depth videos, 91% for IMU data

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Fayez et al. [26]; VaIS	27 (3 F and 24 M)	Visual and Interial dataset focusing on performing the squats exercises	Left Upper Arm, Right Upper Arm, Left Wrist, Right Wrist, Left Calf, Right Calf, Left Thigh, Right Thigh, Lower Back	IMU unit "Meta-Motion" manufactured by "MbientLab"; used only two sensors, the accelerometer and the gyroscope. For visual modality, used the back camera of a regular mobile phone using a 720*p* camera at 30 frames/s	Subjects overview of the dataset (mean (sd)): Height (cm): 170 (15) and 150 (4), and Weight (kg) 80(10) and 60 (5), Age (y): 20 (3) and 25 (4)
Biagetti et al.[264]	7 (4 F, 3 M)	5 sessions each of squat exercises, stepper exercises, resting	Wrist	Maxim Integrated MAXREFDES100 device; Photoplethysmography (PPG) signal and 3-axial accelerometer	Age range 20–52 years; 31.5714 ± 13.6120 years, height range 1.57–1.80 m; weight range 55–83 kg. BMI 23.5429 ± 2.5310 kg/m^2. The dataset recording total duration is 17201 s
Lee et al. [265]	39	Squat posture	Left thigh, right thigh, left calf, right calf, lumbar	5 IMUs; Accelerometer, gyroscope	Each subject made 6 repetitions of acceptable squat, 5 incorrect squats. Deep learning overall accuracy; 91.7%
Tahir et al. [68]; IMSB	20	6 sporting behaviors; football, skipping, basketball, badminton, cycling, table tennis	At the knee, below the neck, wrist regions	3 MPU-9250 IMU sensors; 3-axial values of accelerometers, gyroscopes, magnetometers	Age range 20–30 years and weight range 60–100 kgs
Stromback et al. [31]; MM-Fit	10	Training exercises; squats, lunges (with dumbbells), bicep curls (alternating arms), sit-ups, push-ups, sitting overhead dumbbell triceps extensions, standing dumbbell rows, jumping jacks, sitting dumbbell shoulder press, and dumbbell lateral shoulder raises	2 smartwatches, one on each wrist, an earbud in left ear, and 2 smartphones, one in left trouser pocket (Huawei P20) and one in right trouser pocket (Samsung S7)	Orbbec Astra Pro (RGB, Depth); Mobvoi TicWatch Pro (Accelerometer, Gyroscope, Heart beats per minute); eSense (Accelerometer, Gyroscope); Samsung S7 (Accelerometer, Gyroscope, Magnetometer); Huawei P20 (Accelerometer, Gyroscope, Magnetometer)	Subjects had access to 2 dumbbells with adjustable weight of up to 7.5 kg each
Tuncer et al. [266]	8 (4 F, 4 M)	Daily sport activity recognition; 19 classes, gender recognition; 2 classes	Torso, right arm, left arm, right leg and left leg units	Triaxial accelerometer, gyroscope, magnometer within 9 Xsens MTx sensors	Age range 20–30 years. 99.61% and 99.96% accuracy rates for daily sport activity recognition and gender recognition, respectively
Zhang et al. [267]	9 (4 F, 5 M)	8 activities; standing, walking, squatting, cleaning up template, fetching/placing rebar, locating rebar, binding rebar, placing concrete pad	Right wrist and right leg	2 iPhone 6 s smartphones; 3-axial accelerometer, 3-axial gyroscope	Data collection; frequency 10 Hz, collection times 60 s. CART algorithm; decision tree; tenfold cross validation; overall accuracy 89.85%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Calvo et al. [268]; HAD	16	5 activities; stand still, squat and stand up, jump, raise right hand, jog	4 IMUs located near the wrists and knees. 4 EMGs tracking the activity of 4 body muscles; 2 arms and 2 thighs	4 IMUs, 4 electromyographic sensors (EMGs), Microsoft Kinect visual depth sensors	Using SVM and HMM with Kinect + IMUs + EMG average accuracy $98.81 \pm 1.81\%$
Koskimaki et al. [269]; MyoGym	10	30 different gym exercises each consisting of 10 repetitions	Myo placed at right forearm; IMU placed on top of forearm, 8 electromyography sensors evenly distributed around arm	Myo Armband; 3-axis accelerometer, 3-axis gyroscope, 8 electromyography sensors	Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) with average accuracy 71.6%
Liu et al. [270]		3 groups of activities; group A; lateral arms stretching, walking away, waving hand. group B; longitudinal arms stretching, sideways, pushing. group C; dancing with stretching arms, running away slowly, waving a fist		Pyroelectric infrared (PIR) binary sensor nodes and arrays	CNN recognition accuracy for "standard activities" only 99.5%, for all test samples 97.7%
Lim et al. [271]	3 (3 M)	Walk, Slip, Trip	Left hip pocket	Smartphone; 3-axial accelerometer	Average age 25 years, average weight 72 kg, average height 172 cm. Artificial neural networks (ANN); overall accuracy 94.00%
Chen et al. [272]; UTD-MHAD	8 (4 F, 4 M)	27 activities; right arm swipe to left/right, right hand wave, 2 hand front clap, right arm throw, cross arms in chest, basketball shoot, right hand draw x or circle (clockwise/counterclockwise), draw triangle, bowling (right hand), front boxing, baseball swing from right, tennis right hand forehand swing, arm curl (2 arms), tennis serve, 2-hand push, right hand knock on door/catch an object/pick up and throw, jogging/walking in place, sit to stand, stand to sit, forward lunge (left foot forward), squat (2 arms stretch out)	Kinect camera placed on a tripod 3 ms in front of the participant for his/her entire body to appear in the camera field-of-view. Wearable inertial sensor: right wrist or right thigh	1 Microsoft Kinect camera (RGB videos, depth videos, skeleton joint positions) and 1 wearable inertial sensor (3-axis accelerometer and 3-axis gyroscope)	Every participant performed each activity 4 times. Collaborative Representation Classifier (CRC) using both Kinect and Inertial Sensors; overall accuracy 79.1%.

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Barshan et al. [102]	8 (4 F, 4 M)	19 activities; sitting, standing, lying on back and right side, ascending and descending stairs, standing still/moving around in elevator, walking in parking lot, walking on treadmill flat/inclined, running on treadmill, exercising on stepper, exercising on cross trainer, cycling on exercise bike horizontal/vertical, rowing, jumping, playing basketball	5 different body locations; right-hand side of right knee, left-hand side of left knee, chest, right/left wrists	5 MTx 3-DOF orientation trackers, Xsens Technologies; each unit has 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer	Each activity is performed for 5 min. age range 20–30 years. Classification accuracy; ANNs (99.2%), SVMs (99.2%), GMM (99.1%)
Mitchell et al. [273]	15	Sporting activities; stationary (0) m/s, walking (1 ± 1) m/s, jogging (3.5 ± 1.5) m/s, sprinting (5+) m/s, hitting ball, attempting standing tackle, dribbling ball	Upper crevice of back	Smartphones; Google Nexus One (3-axial accelerometer), HTC Desire (3-axial accelerometer)	SVM approach, optimized classification model, fusion of classifiers; Fusion of classifiers achieved maximum F-measure accuracy 87%
Banos et al. [13, 21]; REALDISP dataset	17	33 different activities; focuses on warm up, fitness, cool down exercises; sensors placed on 9 obtrusive body parts. Activities: Walking, Jogging, Running, Jump, Trunk twist, Waist/Lateral bends, Forward stretching, Arms elevation/ crossing, Cycling, Hand claps, Knees bending, Shoulders rotation and Rowing	Left and right calf, Left and right thigh, Left and right lower/upper arm, Back	3D acceleration, gyroscope, magnetic field, and 4D orientation	Called 'Realistic Sensor Displacement' (REALDISP) dataset. Originally consisting of 4000+ varying-length samples, 1397 of these were found to be usable after preprocessing and the subsequent elimination of indeterminate/unlabeled samples. Xsens IMU units were used to capture the data at a sampling rate of 50 Hz. Samples are recorded while the data capture device was displaced in one of three ways, which was done to capture real-world placement scenarios. See Fig. 3b for data collection setup
Bayati et al. [274]; Fitness activity dataset	1 (0 F, 1 M)	5 aerobic movements of the leg	5 sensors placed on lower leg, 5 sensors on thigh	10 bluetooth acceleration sensors	Expectation Maximization + static LDA; average accuracy is about 62% (sensors on thigh) and 76% (sensors on lower leg)
Liu et al. [275]	50 (31 F, 19 M)	13 activities; computer work, filing papers, vacuuming, moving boxes, self-paced walk, cycling with 1-kp resistance, cycling with 2-kp resistance, treadmill at 3.0 mph, treadmill at 4.0 mph 0% grade, treadmill at 4.0 mph 5% grade, treadmill at 6.0 mph 0% grade, basketball, tennis	Accelerometers (hip and wrist), ventilation sensor attached to the abdomen	Two 3-axial MMA7260QT accelerometers and one ventilation sensor	Age 32.6 ± 9.9 years, weight 67.7 ± 12.3 kg, height 171.2 ± 8.6 cm, BMI 23.2 ± 4.6 kg/m^2. Each activity is recorded for 7 min and followed by 2 min rest. SVM with "leave-one-subject-out" cross-validation procedure; overall accuracy 88.1%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Altun et al. [57] ; Daily and sports activities	8 (4 F, 4 M)	19 activities; postures, motions	5 sensor units worn on chest, arms, legs	5 Xsens IMU units; 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer	9120 fixed-length samples; sampling rate of 25 Hz; BDM: 99.2%, SVM: 98.8%, 7-NN: 98.7%, DTW: 98.5%, ANN: 96.2%, LSM: 89.6%, RBA: 84.5%
Tunçel et al. [276]	1 (0 F, 1 M)	8 motions; standing without moving legs, moving right leg lower part only to back, moving right leg lower and upper parts to front with bending knee, moving right leg forward/backward without bending knee, opening right leg to body right side without bending knee, squatting moving both upper and lower leg parts, moving right leg lower part only upwards while sitting on seat	Right leg	Two 1-axial Gyrostar ENV-05A piezoelectric vibratory gyroscopes	Total signal recording per leg motion is 576 s. Leave-One-Out (LOO) cross validation, PCA applied to 14 features (6 features); Overall accuracy; Bayesian Decision Making (BDM): 99.3%, SVM: 98.4%, 1-NN: 98.2%, DTW-2: 97.8%, LSM: 97.3%, ANN: 88.8%
Ermes et al. [277]	12 (2 F, 10 M)	9 activities; lying down, sitting and standing, walking, running, Nordic walking, rowing with rowing machine, cycling with exercise bike, cycling with real bike, playing football	Hip and wrist	2 3-axial accelerometers, GPS	Age range 27.1 $\pm$ 9.2 years, BMI 23.8 $\pm$ 1.9. Activity data is recorded for 68 h. Leave-one-subject-out cross-validation. Hybrid model (tree structure of a priori knowledge + ANN): overall accuracy 89%, ANN: 87%, custom Decision Tree: 83%, automatic Decision Tree: 60%
Tapia et al. [278]	21	30 physical gymnasium activities; lying down, standing, sitting/fidget feet legs/fidget hands arms, walking (treadmill speed; 2, 3, 4 mph and inclination; 4, 8, 12 degrees), running 5 mph, ascend stairs, descend stairs, cycling (speed; 60, 80, 100 rpm and resistance level; light, moderate, hard), rowing (speed; 30 spm and resistance; light, moderate, hard), carrying weight 2 mph, moving weight high/low/side, calisthenics bicep curls/jumping jacks/push ups/sit ups	Dominant wrist, ankle side, upper arm outside part, thigh upper part, hip; heart rate monitor on chest	Five 3-axial accelerometers, heart rate monitor	Age range 18–65 years. Each activities 2 min, total 16.6 h annotated data are recorded. C4.5 Decision Tree (DT) classifier subject-dependent training; tenfold cross-validation: average accuracy 94.9%. subject-independent training; leave-one-subject-out: average accuracy 56.3%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Olguın and Pentland [279]	3	8 activities; sitting down, running, squatting, walking, standing, crawling, laying down on chest, hand movement while standing	Right wrist, left hip, chest	Three 3-axial accelerometers	Data recorded for each activity 90 s. HMM using K-fold cross-validation; overall accuracy 92.13%
Chambers et al. [280]	1	3 Kung Fu martial art movements; Cuts (wood/grass/throat/side), elbows, punch blocks	Wrist	One 2-axial ADXL202 accelerometer	Multi-layer HMM, overall accuracy 96.67%
Herren et al. [281]	20 (10 F, 10 M)	Running speed and incline (in free-living conditions)	3 accelerometers at lower back, 1 at heel	Four 3-axial accelerometers	Every user ran 18 runs outdoors at various speeds/inclines. About 2 h of recording. Neural Networks; speed Root Mean Square Error (RMSE) is 0.12 m/s, incline RMSE is 0.0142 radiant
Worker activity recognition					
Xia et al. [282]	6	Place board on workbench, remove film from board, screw in parts, pack accumulated boards into box once 20 are completed	Worker's right wrist	Sony SmartWatch3 SWR50; 3-axial accelerometer	Average F-measure across all six work processes 0.81
Akhavian et al. [283]	4 (2 F, 2 M)	7 activities performed on wood sections in outdoor experimental area; cutting lumber (1 activity; sawing), transportation (4 activities; loading, pushing, unloading, returning), installation (2 activities; hammering and wrench turning)	Smartphones fixed at workers' upper arms	Smartphone; accelerometer and gyroscope	Stroboscope (STate and ResOurce Based Simulation of COnstruction ProcEsses); Discrete Event Simulation (DES) model; Activity Cycle Diagrams (ACDs). Neural Network, Decision Tree, k-Nearest Neighbor (k-NN), Logistic Regression, SVM; tenfold stratified cross validation; bagging classification accuracy; cutting lumber 99.28%, transportation 90.09%, installation 92.97%
Activity daily living (ADL) activities					
Li and Wang [105]	5	6 activities: sitting, standing, walking, running, going upstairs, going downstairs	Under the subject knee; can obtain movement information of subject body's lower limbs	9-axial MPU9250 IMU; 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer	Age range 22–32. Dataset contains 130,056 samples
Chifu et al. [284]; ADLS (Activity of daily living set)	10 (5 F, 5 M)	ADLs; sleeping, eating, personal hygiene, reading, spare time/TV, walking, outside	Beacons distributed in rooms	Bluetooth Beacons; detect presence of elderly adults wearing smartwatch	k-fold cross-validation. Markov model-based method. confidence of activity duration 0.98
Zhang et al. [285]	10 (5 F, 5 M)	4 activities; sitting, walking, walking upstairs, walking downstairs	Worn on wrist	3-axial accelerometer, 3-axial gyroscope, temperature sensor	ADL Recognition COVID-19 Prevention. age range 21–60 years, height range 163–180 cm, weight range 45.7–100.1 Kg. Random forest (RF) is the best classifier, overall accuracy 88.78%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Lakoju et al. [286]	2 (1 F, 1 M)	6 activities; dropping/picking up, standing, walking, sitting, stationary on side table, vibrating surface	Subjects held the device	Apple iPad device; 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer, orientation (azimuth, pitch, roll)	Total of 148,778 samples. Fuzzy C-means algorithm achieves the highest accuracy; F-measure 87% and Matthews correlation coefficient (MCC) 84%
De Sousa et al. [287]	9 (2 F, 7 M)	5 ADL activities (walking, running, sitting, standing up, bending to pick an object) and fall activities	A belt; one prototype on back, one in front of body	3-axial LIS3DH accelerometer	Height average 174.11 ± 6.98 cm, weight average 63.57 ± 7.91 kg. Total 135 falls and 86 ADLs were recorded. threshold-based algorithm; false alarms specificity 97.7% and sensitivity 92.6%
Skoglund et al. [288]	21	9 activities; standing, sitting, lying face-up, lying face-down, lying side, walking, jogging, falling, transitioning	2 XSens MTw Awinda at ear-level and 1 at waist-level	3 tri-axial accelerometers	Age range 24–60 years, height range 1.55–1.95 m. Data per subject is about 13–14 min. Combination of Bagging and Classification tree methods; overall accuracy is 91% (ear-level + waist-level)
Gochoo et al. [289]; IM-WSHA	10 (5 F, 5 M)	11 ADLs, brushing hair, cooking, drinking, exercise, ironing, phone conversation, reading book, using computer, vacuum cleaning, watching TV, walking	Wrist, thigh, chest	3 body-worn IMU sensors (MPU-9250): 3-axial accelerometers, gyroscopes, magnetometers	Stochastic Gradient Descent (SGD) used to select optimized features. recognition accuracy; 83.18%, 94.16%, 92.50% using IM-WSHA, PAMAP2, HuGaDB datasets
Sikder et al. [123]; KU-HAR	90 (15 F, 75 M)	18 activities; stand, sit, talk-sit, talk-stand, stand-sit, lay, lay-stand, pick, jump, push-up, sit-up, walk, walk-backward, walk-circle, run, stair-up, stair-down, table-tennis	A waist bag to hold smartphone	5 Android smartphones; Samsung Galaxy J7 (2017), Xiaomi Redmi Note 4, Realme 3 Pro, Realme 5i, and Realme C3. Each smartphone has a tri-axial accelerometer and a tri-axial gyroscope	Age range 18–34 years, age average 21.7 years, weight range 42.2–100.1 kg, weight average 63.2 kg, Two subjects with heart conditions
Choudhury et al. [290]; HAR Sense	12	6 ADLs; running, walking, standing, sitting, upstairs, downstairs	A lumbar belt for waist area and smartphone was kept in the front pockets vertically	Smartphone used: Poco X2 and Samsung Galaxy A32s; Inbuilt sensors used: Accelerometer and Gyroscope	All subjects are above 23 years and above 50 kgs
Palmerini et al. [291]	15 (8 F, 7 M)	15 ADL recordings × 12 h per subject	On the lower back	3-axial accelerometer	Total of 180 h of ADL recordings. age range 68.1 ± 15 years. Subject-based cross-validation
Li et al. [292]	4 (2 F, 2 M)	8 ADLs: shaking hands, waving hand, lying, walking, picking things up, bending over to pick things up, sit-down-stand-up, open-close door	Wrists	3 tri-axial accelerometers	Weighted ensemble method of 5 base classifiers; 3 k-NN with three feature extractors; polynomial curve fitting (PCF), discrete Fourier transform (DFT), piece-wise local statistics (PLS)/Attention-LSTM/CNN; achieves the highest overall accuracy 95.58%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Bhat et al. [293]; w-HAR	22 (8 F, 14 M)	7 activities; jump, lie down, sit, stand, stairs down, stairs up, walk	IMU on right ankle, stretch sensor is sewed to knee sleeve	Accelerometer, gyroscope, stretch sensor	Age range 20–45 years, height range 150–180 cm. w-HAR total duration is about 3 h. Neural network; overall accuracy 95%
Wang et al. [294]; Sussex–Huawei Locomotion (SHL)	3	8 locomotion and transportation activities: still, walk, run, bike, bus, car, train, subway	The training data; 4 smartphones at 4 body positions: hand, torso, bag, hip pocket. The testing position is "Hips"	7 sensors; accelerometer, gyroscope, magnetometer, linear acceleration, gravity, orientation, ambient pressure	2812 h of labeled data, 16,732 km travel distance, CNN best submission achieved F1 score 88.5%
Garcia et al. [295]	19	4 activities: inactive, active, walking, driving	Each subject has his own way of using his smartphone	Smartphone; 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer, GPS (latitude, longitude, altitude, speed, bearing, accuracy of every measurement)	Support vector machine (SVM) mean accuracy of $69.28\% \pm 15.10$ using accelerometer + gyroscope + magnetometer + GPS, mean accuracy $74.39\% \pm 10.75$ using accelerometer + magnetometer + GPS
Ashry et al. [129]; CHAR-SW	25 (14 F and 11 M)	Streams of activities of daily living. Individual actions are typical ADLs including eating, sporting, driving, hygiene, etc	IMU on-board Apple Watch Series 4 on their right wrists	Acceleration, angular velocity, rotation displacement and gravity readings, all of which are tri-axial, recorded at a sampling rate of 50 Hz	Figure 18a shows a snapshot streaming actions
Ebner et al. [296]	8	4 activities; walking, standing, stairs up, stairs down	Hand-carried smartphones	3-axial accelerometer, 3-axial gyroscope	PCA and SVM classifier with the RBF kernel; overall accuracy of 98.0% for pocket-case and 81.8% for hand-case; trained on a combination of MobiAct dataset [150] and this dataset [296]
Alves et al. [297]; Dataset 2	22 (5 F, 17 M)	Standing still, sitting, walking, running, standing with freedom of movements, laying	Chest, waist, left/right frontal pockets	Three-four 3-axial accelerometers	Age average 26 ± 3 years. 272 falls (44.67 h). state machine algorithm; fall detection rate 86.8%
Weiss et al. [298]	51	18 ADLs: walking, jogging, stairs (ascending and descending), sitting, standing, kicking soccer ball, dribbling basketball, catch tennis ball, typing, writing, clapping, brushing teeth, folding clothes, eating pasta/soup/sandwich/chips, drinking from cup	Smartphone in right pants pocket and smartwatch on dominant hand	Accelerometer and gyroscope; Google Nexus 5/5X or Samsung Galaxy S5 smartphone with operating system Android 6.0 (Marshmallow) and LG G Watch with operating system Android Wear 1.5	UCI WISDM HAR and biometrics dataset. age range 18–25. Each activity is 3 min. Random Forest activity recognition average accuracy of 94.4%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Balli et al. [299]	5 (2 F, 3 M)	8 ADLs; brushing teeth, walking, running, vacuuming, writing on board, writing on paper, using keyboard, stationary state	Smartwatch worn on wrist	Accelerometer, gyroscope, step counter, heart rate sensors of Moto 360 smartwatch	HAR from smartwatch using PCA and RF. age range 25–33 years, height range 160–178 cm, weight range 48–95 kg
Reiss et al. [300]; PPG-DaLiA	15 (8 F, 7 M)	8 ADLs; sitting still, ascending/descending stairs, table soccer, cycling, driving car, lunch break, walking, working	RespiBAN Professional chest-worn, Empatica E4 wrist-worn	RespiBAN Professional has 3 modalities: electrocardiogram, respiration, 3-axis accelerometer. Empatica E4 has 4 modalities: photoplethysmogram, 3-axis accelerometer, electrodermal activity, body temperature	36 h of recording; age range 21–55 years; (mean and variance: 30.60 ± 9.59 years)
Ni et al. [301]	10 (3 F, 7 M)	12 activities; standing, sleeping, watching TV, walking, running, sweeping, stand-to-sit, sit-to-stand, stand-to-walk, walk-to-stand, lie-to-sit, sit-to-lie	Wrist-belt attached at left wrist	3-axial accelerometer	Dataset is recorded over 10 h. Ensemble model constructed from SVM, DT, k-NN with tenfold cross-validation; overall accuracy 96.87%
Liu et al. [302]	5 (1 F, 4 M)	4 housekeeping activities; cleaning window (CW), cleaning table (CT), sweeping floor (SF), mopping floor (MF), and the transitions between CW, CT, SF, MF	Dominant wrist	3-axial accelerometer and 3-axial gyroscope	Age average 65.4 ± 3.1 years, height average 156.8 ± 3.2 cm, weight average 58.8 ± 4.1 kg. 948 collected instances. Leave-one-subject-out (LOSO) cross validation approach. best overall accuracy, recall, precision: 81.63%, 78.40%, 78.58%
Ashry et al. [14]; EJUST-ADL-2 (aka HAD-AW)	16 (9 F, 7 M)	Thirty-one activities including eating, driving, cooking actions, house cleaning actions, different kinds of sporting and hobbies	Right wrist only	3D accelerometer, 3D angular velocity, 3D rotation, 3D gravity	Large dataset is collected by only an Apple watch on right wrist. 31 different activities are recorded. Figure 4 shows snapshots of the data collection process
Gomaa et al. [24]; EJUST-ADL-1 (Gomaa-1) dataset	3	Use telephone, Drink from glass, Pour water, Eat with knife/fork, Eat with spoon, Climb/Descend stairs, Walk, Get up/Lie down bed, Stand up/Sit down chair, Brush teeth, Comb hair	Right wrist only	3D accelerometer, 3D angular velocity, 3D rotation, 3D gravity	Fourteen activities are collected by Apple watch on right wrist. Examples of data collection are shown in Fig. 5. Figure 7 shows samples of accelerometer and gyroscope signals of the activity 'walk'.

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Sztyler et al. [151]; RealWorld human activity recognition (RWHAR) dataset	15 (7 F, 8 M)	8 activities: climbing stairs up/down, jumping, lying, standing, sitting, running/jogging, walking	7 locations; the head, chest, upper arm, waist, forearm, thigh, shin	7 wearable smartphone/smartwatches with 6 sensors Accelerometer, Gyroscope, GPS, Light, Magnetometer, Microphone	Age range 31.9 ± 12.4 years, height range 173.1 ± 6.9 cm, weight range 74.1 ± 13.8 kg. Random Forest F-measure; device on-body position 89%, activity subject-specific 84%, activity cross-subjects 79%
Micucci et al. [303]; UniMiB SHAR dataset	30 (24 F, 6 M)	9 types of ADLs (e.g., walking, sitting, or standing) and 8 types of falls (e.g., forward, syncope, or backward)	Place the smartphone in front trouser pockets: half of time in the left pocket and remaining time in right pocket	Samsung Galaxy Nexus I9250 with Android OS version 5.1.1 equipped with Bosh BMA220 acceleration sensor; a tri-axial low-G acceleration sensor	11,771 samples describing both ADLs (7579) and falls (4192), ages range 18–60 years (27 ± 12 years), body mass range 50–82 kg (64.4 ± 9.7 kg), height range 160–190 cm (169 ± 7 cm)
Khowaja et al. [95]; SBA dataset	10 (0 F, 10 M)	7 activities; walking, standing, jogging, sitting, cycling, walking upstairs, walking downstairs	Chest, right wrist, left ankle	4 smartphones and 1 smartwatch; only 1 smartphone and 1 smartwatch are considered; 2 accelerometers, 2 gyroscopes, 1 magnetometer sensor	Error-correcting output coding (ECOC); Hierarchical classification model using QDF and Ensemble classifiers; overall accuracy 95.91%
Weiss et al. [304]	17	18 activities; walking, jogging, ascending stairs, sitting, standing, kicking soccer ball, dribbling basketball, playing catch with tennis ball, typing, handwriting, clapping, brushing teeth, folding clothes, eating (pasta/soup/sandwich/chips), drinking	Smartwatch on dominant hand and smartphone in front-right pocket	LG G Watch smartwatch (accelerometer/gyroscope) and Samsung Galaxy S4 smartphone (accelerometer)	Each activity was recorded for 2 min. Tenfold cross validation. Random Forest classifier using smartwatch accelerometer; overall accuracy; 93.30% (personal model; training data from only the intended subject). 70.30% (impersonal model; training data from all subjects except the intended one)
Shoaib et al. [305]; ComplexHAR	10 (0 F, 10 M)	13 activities; standing, jogging, sitting, biking, typing, writing, walking, upstairs, downstairs, drinking coffee, talking, smoking, eating	Right pocket and right wrist	2 mobile phones (Samsung Galaxy S2); accelerometer and gyroscope	Age range 23–35 years. A dataset consisting of 390 min (30 min for each activity). Naive Bayes, k-NN, Decision Tree; for example, jogging, biking, typing; F-measure is 96%, 93%, 96%, respectively (by Naive Bayes) at 2-s window (accelerometer and gyroscope at wrist)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Twomey et al. [306]; The SPHERE challenge	10	ADLs (e.g., activities like meal preparation, watching television, etc.) and posture/ambulation (e.g., walking, sitting, transitioning, etc.)	A tri-axial accelerometer on wrist, 3 RGB-D cameras installed in-house; living room, hallway, kitchen	Accelerometer, ASUS Xtion PRO RGB-D cameras video with depth information, passive environmental sensors	
Davis et al. [307]; HAR Ambient Assisted Living (HAR AAL)	31	6 basic activities; walking, walking up/downstairs, standing, sitting, laying	A waist-mounted smartphone belt on left/right side	Samsung Galaxy S II smartphones with builtin accelerometer and gyroscope sensors	Age range 22–79 years. 3 machine learning methods; SVM, ANNs, and SVM-HMM achieves best classification accuracy 99.7%
Bruno et al. [249]; Wearable Human Activity Recognition Folder (WHARF)	17 (6 F, 11 M)	14 ADLs; brush own teeth, comb own hair, get up/lie down on bed, sit down/stand up from chair, drink from glass, eat with forkandknife/spoon, pour water into glass, use telephone, climb/descend stairs, walk	Right wrist	3-axial accelerometer	Age range 19–81 years
Banos et al. [72]; Mhealth	10 (5 F, 5 M)	12 outdoor activities; standing still, lying down, sitting and relaxing, walking, jogging, running, cycling, climbing stairs, crouching, waist bends forward, jump front and back, and frontal elevation of arms	Left ankle, the right wrist and the chest	One 2-lead Electrocardiogram (ECG) sensor, two 3-axis gyroscope sensors, three 3-axis magnetometer sensors, and three 3-axis accelerometer sensors	
Piyare et al. [308]	4 (2 F, 2 M)	6 ADLs; walking, running, sitting, walking upstairs, walking downstairs, standing	Front pant pocket	Accelerometer, gyroscope, magnetometer	Age range 25–30 years, weight range 60–85 kg, height range 162–187 cm, BMI range 18.5–29.9 kg/m^2. accuracy up to 99.33%
Lockhart and Weiss [309]	59	6 ADLs; walking, jogging, ascending stairs, descending stairs, sitting, standing, lying down	Pocket	15 Android smartphone models; accelerometers	Ten-fold cross validation. Multilayer Perceptron Neural Network; overall accuracy 98.70% (personal model; training data from only the intended subject). Random Forest classifier; overall accuracy 75.9% (impersonal model; training data from all subjects except the intended one)
Khan et al. [310]	30 (12 F, 18 M)	15 activities; walking, walking treadmill, running, running treadmill, going upstairs, going downstairs, elevator up, elevator down, riding bike, hopping, idle (sitting/standing), watching TV, vacuuming, driving car, riding bus	Trousers front pockets, trousers back pockets, jackets inner pocket	LG NEXUS 4 smartphones; 3-axial accelerometer, pressure sensor, microphone	Age range 26–35 years, average height 172.4 cm, average weight 64 kg. Kernel Discriminant Analysis (KDA)-based SVMs for subject-independent online recognition test; overall accuracy 92.40%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Krishnan et al. [131]	3	11 activities; bathing, bed to toilet, cook, eat, enter home, leave home, other, personal hygiene, relax, sleep, take medicine	Sensor layout in three apartments	Binary motion sensors that are triggered by human motion	SVM
Siirtola and Röning [311]	7	5 activities: walking, running, cycling, driving car, sitting/standing	Trousers front pocket, jacket pocket, backpack, brachium, ear	Nokia N8 smartphone; 3-axial accelerometer	Age range 24–34 with average 29 years and height range 1.65–1.85 with average 1.78 m. Total duration of data recorded is about 15 h. Decision Tree, Quadratic Discriminant Analysis (QDA); offline recognition; leave-one-out method; average accuracy is 98.9%. online study; average accuracy is over 90%
Bruno et al. [16]	16 (5 F, 11 M)	14 ADL activities; brush teeth, climb stairs, comb hair, descend stairs, drink glass, eat meat, eat soup, getup bed, liedown bed, pour water, sitdown chair, standup chair, use telephone, walk	Subject right-wrist	1 tri-axial accelerometer	https://archive.ics.uci.edu/ml/datasets/Dataset+for+ADL+Recognition+with+Wrist-worn+Accelerometer age range 19–83. Gaussian Mixture Modeling and Gaussian Mixture Regression. Dynamic Time Warping and Mahalanobis distance
Cleland et al. [312]	8 (0 F, 8 M)	7 activities; walking, jogging on motorized treadmill, sitting, lying, standing, walking up, down stairs	Chest, lower back, left hip, left thigh, left wrist, foot	6 Shimmer wireless 3-axial accelerometers	Age range 24–33 years; mean 26.25 years, standard deviation ± 2.86 years. SVM, C4.5 DT (J48), NB, NN (Multilayer Perceptron), tenfold cross validation; average accuracy 96.67%, 94.18%, 94.77%, 95.74%, respectively
Prudêncio et al. [313]	1	5 activities; standing, walking, running, upstairs, downstairs	In hand in front of body	LG Maximo Black device; 3-axial accelerometer, orientation sensor	80 samples of 5 s for each activity. ANN; accuracy above 90% for single user and phone
Trabelsi et al. [314]	6	12 activities: stairs down, standing, sitting down, sitting, from sitting to sitting on ground, sitting on ground, lying down, lying, from lying to sitting on ground, standing up, walking, stairs up	Chest, right thigh, left ankle	Three 3-axial accelerometers	Age range 25–30, weight range 55–70 kg. Tenfold cross-validation. Multiple HMM Regression (MHMMR) using Expectation-Maximization (EM); highest unsupervised classification accuracy $91.4\% \pm 1.65\%$. k-NN ($k = 1$); highest supervised classification accuracy $95.8\% \pm 0.32\%$
Thiem-jarus et al. [315]	8 (2 F, 6 M)	6 activities; lying, sitting, standing, walking, running, jumping	Pocket, waist, trouser (left), trouser (right)	iPhone; 3-axial accelerometer	Age range 19–45 years. k-Nearest Neighbor (k-NN), overall accuracy 78.23%
Anguita et al., 2013; Smartlab-CETpD [45]	30	6 ADL activities: walking, walking upstairs, walking downstairs, sitting, standing, laying	Smartphone (Samsung Galaxy S II) worn on the waist	Smartphone: accelerometer and gyroscope; 3-axial linear acceleration and 3-axial angular velocity at 50 Hz	Ages range 19–48 years

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Chavar-riaga et al., 2013; The Opportunity Challenge [146]	4	6 runs per subject; 5 ADLs and a "drill" run to collect large number of activity instances through a scripted sequence of activities. The ADL run consists of situations; e.g., preparing sandwich which has composite activities (e.g., cutting bread) and atomic activities (e.g., reach for bread, move to bread cutter, operate bread cutter)	IMUs mounted on a jacket, accelerometers on the limbs, InertiaCube3 located on each foot	5 XSense IMUs (accelerometer, gyroscope, and magnetic sensors), 12 Bluetooth 3-axis accelerometers, 2 InertiaCube3 sensors	Feature vector has 9 dimensions per IMU mounted on jacket, 3 per accelerometer, 16 per InertiaCube3 sensor
Leutheuser et al. [316]; DaLiAc	23 (10 F, 13 M)	13 activities; postures (sitting, lying, standing), household (washing dishes, vacuuming, sweeping), walking (normal, running, stairs climbing), and sports (bicycling, rope jumping)	Placed on wrist, hip, chest, and ankle	4 sensor units, each consisting of a tri-axial accelerometer and a tri-axial gyroscope	A hierarchical multi-sensor classification of a large set of ADLs; an overall mean classification rate of 89.6%; subjects: age range 27 ± 7 years, body mass index (BMI) 24.0 kg/m^2 $\pm$ 3.5 kg/m^2, mean $\pm$ standard deviation (SD)
Ofli et al. [317]; Berkeley MHAD	12 (5 F, 7 M)	11 activities; jumping, jumping jacks, bending, punching, waving 2 hands/1 hand, clapping, throwing, sit down/stand up, sit down, stand up	Accelerometers measure movement at wrists, ankles, hips	Optical mocap system, 12 Dragonfly2 cameras, 2 Microsoft Kinect cameras, 6 three-axis accelerometers, 4 microphones	Age range 23–30 years except 1 elderly subject. 5 repetitions of each action. 82 min of total recording time. Integrating the mocap, accelerometer, Kinect, video, and audio using Multiple Kernel Learning; recognition accuracy 100%
Anguita et al. [62], UCI HAR	30	6 activities; walking, upstairs, downstairs, standing, sitting, laying	Smartphone worn on waist	Samsung Galaxy S2 smartphone: 3-axial accelerometer and gyroscope	Age range 19–48 years. SVM classifier; overall accuracy 89.3%
Casale et al. [253]	10 (3 F, 7 M)	5 activities; climbing stairs, walking, interacting with environment, standing, working at computer	Worn in a little bag around the waist	BeagleBoard, low-cost single-board computer; acquires audio and video through web-cam and motion data through 3-axial accelerometer	Perform activities for minimum time of 5 min. Total recording of 7 h and 11 min. age range 27–35 years. AdaBoost classifier and Leave-One-User-Out (LOUO) cross-validation; False Rejection Rate (FRR) = 0.0072 and False Acceptance Rate (FAR) = 0.03
Liang et al. [318]	24 (8 F, 16 M)	5 static activities; standing, sitting, lying prone, lying supine, driving. 6 repetitive activities; walking, running, ascending stairs, descending stairs, cycling, jumping	Every subject is assigned a smartphone	HTC G11, Samsung i909; 3-axial accelerometer	Age range 22–35. Hierarchical Recognition algorithm; fivefold cross-validation; average accuracy 89.1%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Ugulino et al. [319]	4 (2 F, 2 M)	5 activities; sitting, sitting down, standing, standing up, walking	Waist, left thigh, right arm, right ankle	4 3-axial ADXL335 accelerometers	Total 8 h of activities, i.e., each subject for 2 h. age range 28–75 years, height range 1.58–1.71 m, weight range 53–83 kg. Ensemble AdaBoost combining 10 decision trees; overall accuracy 99.4%
Lara et al. [255]	8 (1 F, 7 M)	5 activities; sitting, walking, running, ascending, descending	Chest sensor strap	BioHarness^{TM} BT chest sensor strap; 3-axial accelerometer, vital signs (heart rate, respiration rate, breath amplitude, skin temperature, posture/ inclination, electrocardiogram amplitude)	Age range 9–34 years. Weight range 27–95 kg. Height range 1.35–1.88 m. BMI range 20.96–29. Tenfold cross-validation. Additive Logistic Regression (ALR) overall accuracy 95.7%, Bagging using ten J48 classifiers (BJ48) overall accuracy 94.24%, Bagging using ten Bayesian Network classifiers (BBN) overall accuracy 88.33%
Albert et al. [320]	26 (19 F, 7 M)	Walking, standing, sitting, holding, or not wearing the phone	Front pocket	T-mobile G1 phones; 3-axial accelerometer	18 healthy subjects and 8 patients. 13 younger subjects (7 F, 6 M), age range 25.1 ± 3.0 years, 5 older subjects (5 F), age range 53.4 ± 7.4 years. 8 patients; Hoehn and Yahr stage 1–3 (7 F, 1 M), age range 67.0 ± 8.1 years, support vector machines, regularized logistic regression; accuracy 96.1% of activities of healthy subjects, 92.2% of activities of Parkinson's patients
Hattori and Inoue [321]; ALKAN	216	46 activities; walk, sit, bicycle, train sit, recline, train stand, eat sit, bus sit, car, stand, stair down, chat sit, stair up, type, test data, motorbike, walk slow, cook stand, elevator down/5fl, elevator up/5fl, run, walk fast, change cloth, escalator up, slope up, escalator down, chat stand, slope down, darts, bus stand, escalator walk up, stretch, weight, bowling, shinkansen (bullet train), escalator walk down, stair run down/up, radio taiso 1, cut grass, ferry, monorail, play courage, barbeque	19 device positions; left pants pocket, left jacket pocket, right pants pocket, no label, right jacket pocket, right hand, left hand, breast pocket, hand bag, shoulder bag, rucksack, neck strap, belt, right wrist, left wrist, left arm, right arm, right ankle, left ankle	216 iPod touches as mobile sensors; 3-axial accelerometer and GPS coordinates	35,310 activity data points. Recursive partitioning tree, Naive Bayes, k-NN, SVM + threefold cross validation. Adopting "left pants pocket", 40 subjects, 8 activities; overall accuracy about 35%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Dernbach et al. [322]	10	Simple activities; biking, climbing stairs, driving, lying, running, sitting, standing, walking, phone not on person. complex activities; cleaning, cooking, medication, sweeping, washing hands, watering plants	Not standard for location/ orientation of smartphone; left to subject choice	Samsung Captivate^{TM} smart phone; 3-axial accelerometer, 3-axial gyroscope	
Schindhelm [323]	1 (1 M)	Standing, walking, jogging, climbing stairs (downstairs and upstairs), taking elevator	3 variations: device into pants pocket, into belt-bag carried by subject, holding device in hand while looking at screen	HTC hero smartphone; BMA150 accelerometer	24 years old male. Decision tree; overall accuracy 92.30%
Reiss et al. [94]; Physical Activity Monitoring for Aging People PAMAP2	9 (1 F, 8 M)	18: 12 activities (lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, ascend, descend stairs), 6 optional activities (watch TV, computer work, drive car, fold laundry, clean house, play soccer)	Chest: 1 IMU and heart rate chest strap. Wrist: 1 IMU on the dominant arm. Ankle: 1 IMU on the dominant side's ankle	3 Colibri wireless IMUs from Trivisio (3-axis MEMS sensors; two accelerometers, gyroscope, magnetometer). BM-CS5SR heart rate monitor from BM innovations GmbH	Age range 27.22 ± 3.31 years, BMI range 25.11 ± 2.62 kg m^{-2}. Total recording of 10+ h from 18 different actions. k-NN classifier with subject independent cross-validation; all activity recognition accuracy 89.24%
Chuang et al. [324]	10	7 activities; sitting, standing, walking upstairs, walking downstairs, walking at normal pace on sports field, running, cycling	Wrist and ankle	3-axial accelerometer	k-NN and Probabilistic Neural Network (PNN) with overall accuracy 96% Each activity is performed for 3 min. Age range 20–25 years, BMI average 21.78 ± 2.82 kg/m^2
Ioana-Iuliana and Rodica-Elena [325]	4 (1 F, 3 M)	8 daily movements; sitting, standing, kneeling, lying with face down, lying with face up, lying on one side, crawling, walking	Right part of hip, lower part of right leg	Two 3-axial accelerometers	Age range 23–27 years. Neural Network, overall accuracy 99.8%
Chern-bumroong et al. [326]	7 (2 F, 5 M)	5 ADLs; sitting, standing, lying, walking, running	Non-dominant wrist	eZ430-Chronos watch; 3-axial accelerometer	Age range $27 - 35$ years. Total data recorded is 35 min. Fivefold cross-validation. Decision Tree C4.5 and Neural Network, overall accuracy 94.13% and 90.45%, respectively
Wang et al. [327]	20	11 ADLs; lying, turn-over, lie-sit, sit-lie, sit-stand/stand-sit, sit-stand, stand-sit, sitting, standing, walking, running	Waist	One 3-axial LIS302DL accelerometer	Age range 20–28 years. Each activity recorded for at least 4 s. Threshold-based, overall accuracy 94.1%
Gjoreski et al. [328]	11 (4 F, 7 M)	7 postures; standing, sitting, lying, standing up, going down (sitting down/lying down/falling), falling on all fours, sitting on ground	Chest, waist, right thigh, right ankle	Four 3-axial accelerometers	Test scenario is about 15 min including all postures. Leave-one-person-out technique. Naïve Bayes, SVM, J48, Random Forest achieves the best overall accuracy 99.0%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Zhang et al. [329]	10 (5 F, 5 M)	postures (sitting, standing, lying), locomotion (walking), transitions (sit-to-stand and stand-to-sit)	Chest and both thighs	Three Functional Activity Monitor (FAM) sensors; 3-axial accelerometer and 3-axial gyroscope	Age range 23 ± 3 years. All postures and locomotion recorded at least 10 s, transitions 2–3 s. Threshold-based, sensitivity at least 96.2%
Jun-Ki and Sung-Bae [330]	3	11 activities; walking, hands-clapping, hands-shaking, teeth-brushing, drinking, keyboard-typing, mouse-scrolling, door-opening, calling, scissoring, writing. 8 composite behaviors; working, meeting, eating, exercising, sleeping, calling, going-to-toilet, resting	Forehead, both arms, both wrists	5 XSens sensor modules (3-axial accelerometers, 3-axial gyroscopes), 2 data-gloves (primitive motion detection), 1 armband sensor (composite behavior detection)	Leave-one-person out cross validation (threefold cross validation). Localized model of SVM and dynamic selection, average accuracy 99.4% (primitive motion detection), 95.99% (composite behavior detection)
Xu et al. [331]	3	10 postures; typing, sit with arms to front, sit with arms up, sit with hands on hip, sit with resting chin, stand, stand with arms to front, stand with arms up, stand with hands on hip, bend forward	1 at waist, two on each arm, two on each leg	Nine 3-axial accelerometers	Every subject performs each posture for 30 s 3 times. Threshold-based, overall accuracy 96.6%
Atallah et al. [332]	11 (2 F, 9 M)	15 activities; lying down, preparing food, eating/drinking, socialising, reading, dressing, walking in corridor, treadmill walking 2 km/h, vacuuming, wiping tables, running in corridor, treadmill running 7 km/h, cycling, sitting down/getting up for 5 times, lying down/getting up for 5 times	Ear, chest, arm, wrist, waist, knee, ankle	7 3-axial ADXL330 accelerometers	Each activity is performed for 2 min. Feature selection algorithms (Relief, Simba, mRMR) + k-NN; this work is studying sensor positioning and feature selection
Andreu et al. [333]	1	8 activities; walking, running, bicycling, standing, laying down, vacuuming, stretching, sitting	SunSPOT sensors (palm, arm), Nokia N97 (pocket, bag, mobile belt bag, etc.), Porcupine sensor (wrist)	SunSPOT sensors (3-axial accelerometer), Nokia N97 (3-axial accelerometer), Porcupine sensor (3-axial accelerometer)	eClass_HAR; recursive PCA and LDA + evolving Fuzzy Rule-Based (FRB) classifier, overall accuracy 71.4%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Martín et al. [334]	5 (0 F, 5 M)	15 activities; walk slowly along corridor, run in same floor along 3 corridors, wait for lift, walk into lift, take lift to go to lower floor, walk out of lift to notice board, stand in front of notice board, go up stairs 2 floors, walk along 2 corridors, go down stairs one floor, walk into office, sit, walk to another office, lie down, get up and stand	Hip, foot, chest	Bio chest belt, three 3-axial accelerometers, oxymeter	Threshold-based, overall accuracy 89.4%
Riboni and Bettini [335]; COSAR	6 (3 F, 3 M)	10 activities; brushing teeth, hiking up/down, riding bicycle, jogging, standing still, strolling, walking downstairs/upstairs, writing on blackboard	HTC Magic device in left pocket, Sun SPOT at right wrist	HTC Magic device: 3-axial accelerometer, Sun SPOT: 3-axial accelerometer, GPS receiver tracks current location	Age range 30–60. Each activity is performed 30 min; activity dataset recorded for 5 h. Fourfold cross validation. Combined Ontological/Statistical Activity Recognition (COSAR) historical variant; overall accuracy 93.44%
Maekawa and Watanabe [336]	40 (19 F, 21 M)	14 activities; stand, walk, run, sit, ascend stairs, descend stairs, bicycle, brush teeth, wash dishes, use PC, draw on whiteboard, write in notebook, play ping pong, vacuum	Wrists of both hands, waist, right thigh	Four 3-axial accelerometers	Age range 25–60 years, height range 155–180 cm, weight range 45–80 kg. 100 h of sensor data. Leave-one-participant-out cross validation. HMM with GMM, overall accuracy about 98%
Lee et al. [337]	20	6 activities; lying, standing, walking, upstairs, downstairs, driving	Chest	3-axial accelerometer	Age range 22–30 years. ANNs average accuracy; subject-independent 94.43%, subject-dependent 96.61%
Alvarez et al. [338]	1	3 postures: sitting, upright, walking	Centered in subject's back	3-axial accelerometer	Leave-one-out cross validation. Genetic fuzzy finite state machine (GFFSM), overall accuracy 98.9%
Reiss and Stricker [339]	8 (1 F, 7 M)	14 activities; lie, sit, stand, iron, vacuum, ascend stairs, descend stairs, soccer, walk very slow, normal walk, Nordic walk, run, cycle, rope jump	Lower arm, chest, foot	Three 3-axial accelerometers and heart rate monitor	Age 27.88 ± 2.17 years, BMI 23.68 ± 4.13 kg/m^2. About 8 h of data are recorded. Leave-one-subject-out eightfold cross-validation protocol. Boosted Decision Tree achieved best results, overall accuracy 90.65%
Kawaguchi et al. [340]; Human Activity Sensing Consortium (HASC) Challenge	540 (12 F, 89 M, 439 Unknown)	6791 activity meta-data files; Females (341), Males (4032), Unknown (2418). 20 teams; every participant gathers at least 5 subjects: 5 sets of 6 activities: stay, walk, jogging, skip, stair-up, stair-down	Various positions, e.g., waist pocket	Every participant use any sensor type available in the market. The majority of sensor types iPhone/iPod Touch and WAA-series (ATR); 3-axis accelerometer	Weight range 40–110 kg, height range 148–189 cm. Recognition results range 38–72%
Gao et al. [341]	8	8 scripted normal ADL, three times each	Waist, chest, thigh, side of the body	4 accelerometers	Naive Bayes, overall accuracy 97.7%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Lee and Cho [342]	3	7 activities; standing, walking, stairs up/down, running, shopping, taking bus, moving by walking	Hold smartphone by hand	Android OS-based HTC Desire; 3-axial accelerometer	Age range 20–30 years. Total recorded data about 5.65 h. Hierarchical HMM; fourfold cross validation, average precision about 83%
Hong et al. [343]; ConaMSN	1 (1 M)	7 activities; sleeping, urinating, calling, working, meeting, resting, exercising	Head, right/left arms, right/left hands	five 3-axial accelerometers, 3-axial gyroscope, 3-axial magnetometer	Total recorded 17,939 samples. Fivefold cross validation. Dynamic Bayesian networks; average recognition rate about 84%
Khan et al. [15]	6 (3 F, 3 M)	15 activities; postures (lying, sitting, standing), motions (walking, walking-upstairs, walking-downstairs, running), transitions (lie-stand, stand-lie, lie-sit, sit-lie, sit-stand, stand-sit, walk-stand, stand-stand)	Chest	One 3-axial MMA7260Q accelerometer	Age mean is 27 years. Sixfold cross validation. Hierarchical-recognition Scheme (LDA + ANN); average accuracy is 97.9%
Lu et al. [344]	16 (4 F, 12 M)	Cycling, vehicle, running, stationary, walking	Each participant carries multiple phones in different body positions; e.g., in front pocket, backpack	Nokia N95; 3-axial accelerometer	Decision Tree, Multivariate Gaussian Model, SVM, Naive Bayes overall accuracy 93.87%
Frank et al. [345], DLR	16 (6 F, 10 M)	7 activities; sitting, standing, walking, running, jumping, falling, lying	Worn on belt	One IMU; 3-axial gyroscope, 3-axial accelerometer	Age range 23–50 years. recorded data 270 min. Fourfold cross validation. HMM based on learnt Bayesian Network (BN); recall rate range 93–100%
Van Kasteren et al. [346]	1 (1 M)	Brushing teeth, showering, bathing, shaving, preparing breakfast, preparing dinner, snacking, drinking	Six-room house; sensors at bathroom door, toilet flush, bathtub passive infrared, dresser passive infrared, microwave, refrigerator, freezer, cupboard with plates/cups/pans/boxes, cutlery drawer	RFM DM 1810	57 years old male. 19 days of sensor data; Hidden semi-Markov models (HSMM)
Khan et al. [347]	10 (3 F, 7 M)	6 ADLs: resting (sit/stand), walking, walk-upstairs, walk-downstairs, running, cycling	Chest pocket, front trousers pocket, rear trousers pocket, inner jacket pocket	3-axial FreeScale MMA7260Q accelerometer	Age mean 45 years, standard deviation 5 years. About 34 h of activity data. average accuracy about 95%
Sazonov et al. [348]	9 (6 F, 3 M)	6 activities; sitting, standing, walking/jogging, ascending stairs, descending stairs, cycling	Shoe back	3-axial LIS3L02AS4 accelerometer	Age range 18–31 years, weight range 55.6–100.9 kg, BMI range 18.1–39.4 kg/m^2. Total 11 h and 36 min are recorded. SVM with fourfold cross validation, highest recognition accuracy 98.1% ± 2.3%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Gu et al. [349]	2	21 activities; single-user ADLs; brushing teeth, washing face, brushing hair, making pasta, making coffee, toileting, ironing, making tea, vacuuming, using phone, using computer, reading book/magazine, watching TV, eating meal, drinking. multi-user ADLs; making pasta, cleaning dining table, making coffee, toileting (with conflict), watching TV, using computer	iMote2 sets and RFID readers on both hands, audio recorder on trapezius	2 Crossbow iMote2 sets (3-axial accelerometer each), 2 RFID wristband readers, 1 audio recorder (human voice and environmental sound)	Final-year university students; collecting real-world activity traces during 2 weeks in smart home environment; 420 annotated instances for both subjects. Emerging Pattern-based Multi-user Activity Recognizer (epMAR); tenfold cross-validation; average accuracy 89.72%
Roy et al. [350]	10 hemiparetic patients (5 F, 5 M)	11 ADLs (identification tasks); food cutting/lifting, tooth brushing, hair combing, shirt buttoning, pant lifting, sit-to-stand, supine-to-sit, sitting, walking, bowel movement. 10 activities (non-identification tasks) to evaluate mis-classification errors; typing, box waist-to-head, open/close drawer, phone call, flipping pages, box floor-to-waist, organizing small objects, writing, folding clothes, standing	Both upper arms, one forearm, one thigh	4 pairs: accelerometer sensors and adjacent surface electromyographic (sEMG) sensors	Age range 33–67 years (mean age 51.7 ± 11.4 years). Multilayered neural network and adaptive neuro-fuzzy classifier; identification tasks; mean sensitivity 95.0% and mean specificity 99.7%. non-identification tasks: mean misclassification error $< 10\%$
Bonomi et al. [351]	20 (7 F, 13 M)	20 activities; lying on bed, sitting on chair, sitting working on computer, standing, washing dishes, walking along corridor, walking downstairs, walking upstairs, walking outdoors (4 patterns), running outdoors (4 patterns), cycling (4 patterns)	Lower back	One 3-axial accelerometer	Age 29 ± 6 years, weight 72 ± 9.0 kg, height 1.74 ± 0.09, BMI 23.6 ± 3.2 kg/m^2. Leave-one-subject-out cross-validation algorithm. Decision tree; highest classification accuracy is 93%
Song et al. [352]	4 (2 F, 2 M)	6 activities; sitting down, standing up, lying, walking, running, falling	Torso	3-axial accelerometer	Females average age 32 years and males average age 36 years. overall accuracy 88%
Maguire and Frisby [353]	6 (2 F, 4 M)	8 activities; standing, brushing teeth, up stairs, down stairs, walking, running, vacuuming, sit ups	Dominant thigh	3-axial accelerometer, heart monitor	Age range 43–45 years. k-NN, J48; tenfold cross-validation; k-NN ($k = 1$) achieves better overall accuracy 90.07%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Győrbíró et al. [354]	3	6 activities; resting, typing, gesticulating, walking, running, cycling	Dominant wrist, hip, and ankle	In-house MotionBand; accelerometer	Average number of samples per activity is 1776 C4.5, NN
Cook et al. [127]	20	Complete activity sensor dataset; 5 ADL activities; telephone use, hand washing, meal preparation, eating/medication use, cleaning. Erroneous activity dataset; 5 ADL activities as the complete dataset, but with activity skipped/performed incorrectly. Inter-weaved activity dataset; 7 ADL activities; filling medication dispenser, filling out/addressing birthday card, selecting outfit job interview, watching DVD, watering plants, answering phone call, sweeping floor. Multiple resident activity dataset; 2 students lived in the apartment for 8 weeks	Sensors distributed about 1 m apart in three-bedroom smart apartment	Motion sensors, digital sensors for ambient temperature readings, analog sensors for hot water/cold water/stove burner use readings	
Yang et al. [76]; UC Berkeley Wearable Action Recognition Database (WARD)	20 (7 F, 13 M)	13 ADLs; stand, sit, lie down, walk forward, walk left-circle, walk right-circle, turn left, turn right, go upstairs, go downstairs, jog, jump, push wheelchair	Sensors placed at 5 places: 2 wrists (left/right), waist front center, 2 ankles (left/right)	3-axis accelerometer and 2-axis gyroscope	Age range 19–75 years. Distributed Sparsity Classifier (DSC) active sensor rate (ASR) = 91.85% when $\tau_1 = 0.08$
Guenter-berg et al. [355]	3	25 actions; stand to sit (armchair/dining chair), sit to stand (armchair/dining chair), sit to lie, lie to sit, bend and grasp from ground/coffee table, turn clockwise/counterclockwise 90°, look back clockwise/counterclockwise and return, kneeling, move forward 1 step, move one leg beside the other, reach up to cabinet, grasp an object, turn clockwise/counterclockwise 360°, etc	Right/left arms, right/left forearms, right/left thighs, right/left ankles	3-axial accelerometer and 2-axial gyroscope	Distributed continuous action recognition. Each action is performed 10 times. Hidden Markov Models clustering accuracy of each subject is: 92.6%, 82.6%, 83.2% when all subjects trained on one model, and 90%, 94%, 94% when all subjects trained individually

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Yang et al. [356]	7 (4 F, 3 M)	8 activities; standing, sitting, walking, running, vacuuming, scrubbing, brushing teeth, working at a computer	Dominant wrist	3-axial accelerometer	Age range 24.1 ± 1.8 years. Each subject performs the 8 activities for 2 min per activity. ANN average recognition accuracy is 95.24%, k-NN average recognition accuracy is 87.18%
Krishnan and Panchanathan [357]	10	7 activities; walking, sitting, standing, running, bicycling, lying down, climbing stairs	Hip, dominant ankle, non-dominant thigh	Accelerometer	SVM, regularized logistic regression LR, boosted decision stumps (AdaBoost) achieves best accuracy (subject independent) 92.81%
Hong et al. [358]	15	18 ADLs; sitting, standing, walking, lying, running, hand shaking, rope jumping, reading + sitting, brush hair + standing, pushing a shopping cart, phone calling + sitting, taking picture + standing, put on skin conditioner, wiping with cloth + standing, running a vacuum cleaner, put on an umbrella + standing, toothbrush + standing, cutting + standing	Wrist, waist, thigh	Three 3-axial accelerometers and glove-type RFID-based iGrabber	Age average 22.9 years. Each activity is performed for 45 s and repeated twice. Decision tree with overall accuracy 95%
Song et al. [359]	7 (2 F, 5 M)	9 ADLs; sit-to-stand, stand-to-sit, lying, lying-to-stand, stand-to-lying, walking, running, sitting, falling down (left, right, backward, forward)	Worn at left side of waistband	3-axial accelerometer	Age range 28–72 years, height range 157–181 cm, weight range 47–71 kg. Total samples 5773 including 274 cases of falling down. Multi-Layer Perceptron; overall accuracy 95.5%
Yeoh et al. [360]	5	3 body postures (lying, sitting, standing) and 3 walking speed movement activities (walking, jogging, running)	Waist and two thighs	3 MTS310CA 2-axial accelerometers	Ruled-based Heuristic model utilizing Extended Kalman (EK) Filtering algorithm; overall classification accuracy 100%; lying, sitting, standing, walking. overall mean-square error (MSE) $1.76(km/h)^2$; walking, jogging, running
Van Kasteren et al. [361]	1 (1 M)	7 activities; leave house, toileting, showering, sleeping, preparing breakfast, preparing dinner, preparing beverage	Three-room apartment; sensors at doors, cupboards, refrigerator, toilet flush sensor	RFM DM 1810	26 years old male. 28 days of sensor data; 2120 sensor events, 245 activity instances. Hidden Markov model and conditional random fields; time slice accuracy 95.6%, class accuracy 79.4%
Huynh et al. [362]; Darmstadt Daily Routine dataset	1 subject over 16 days period	34 ADLs; preparing food, washing hands, discussing at whiteboard, using a computer, walking (freely/carrying something), standing, sitting (desk activities/phone call/having coffee), eating (lunch/dinner/breakfast), brushing teeth, driving car, etc	2 wearable sensors; one in right hip pocket, one around the dominant (right) wrist	Porcupine sensor platform to record activities. A 3D accelerometer (ADXL330) and PIC microcontroller for features preprocessing	Total 164 h of recordings; discard 28 h for sensor failures. 13 dimensions feature vector. Overall accuracy of 72.7% on the activity dataset

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Lo et al. [363]	8	5 activities; reading, walking slowly, lying down, walking, running	Ear-worn	3-axis accelerometer	Bayesian Framework for Feature Selection (BFFS) and Multivariate Gaussian Bayes classifier; average accuracy of 87%
Maurer et al. [9, 364]	6	6 activities: sitting, standing, walking, ascending stairs, descending stairs, running	Left wrist, belt, necklace, in right trouser pocket, shirt pocket, bag	6 eWatch devices; accelerometer, light sensor	Sensor data recorded for over 290 min. C4.5 Decision Tree classification accuracy is 92.8% (bag) using fivefold cross validation
Pirtti-kangas et al. [365]	13 (4 F, 9 M)	17 activities; clean whiteboard, read newspaper, stand still, sit and relax, sit and watch TV, drinking, brush teeth, lie down, vacuum clean, typing, walking, climb stairs, descend stairs, elevator up, elevator down, running, cycling	Right wrist, left wrist, right thigh, necklace	Four 3-axial accelerometers, heart rate	Age range 22–32. k-NN using fourfold cross validation; recognition accuracy is 90.61%
Allen et al. [366]	6 (4 F, 2 M)	8 activities; 3 postures (sitting, standing, lying) and 5 transitions/motion (stand-to-sit, sit-to-stand, stand-to-lie, lie-to-stand, walking)	Waist	3-axial accelerometer	Age range 80–86 years. Adapted Gaussian Mixture Model (GMM) mean accuracy 92.2%. GMM mean accuracy 91.3%. Rule-based Heuristic system mean accuracy 71.1%
Blum et al. [367]	1 (1 M)	Postures: unknown, lying, sitting, standing, walking, running, biking. Activities: no activity, eating, typing, shaking hands, clapping hands, driving, brushing teeth, doing dishes	Left side of hip, dominant wrist	Two 3-axial accelerometers	Total recorded data; 24 h from 11 sessions. Naïve Bayes classifier (using Gaussian model/mixture of Gaussians); posture classification; overall accuracy 91.5%. activity classification; overall accuracy 78.5%
Tapia et al. [368]; MIT PlaceLab	1	Prepare recipe, load dishes, clean kitchen, laundry, make bed, light cleaning, searching, appliances use, phone talk, etc	Environmental sensors: distributed in the environment. wearable sensors: left arm, right arm, left leg, right leg, hip	6 environmental sensors: movement, object-usage-detection, light, temperature, proximity, electric current sensing. 5 wearable sensors: 3-axis accelerometer, heart rate monitor, ultra-violet radiation, RFID reader wristband, location beacons	Small number of subjects. activities definition is unclear

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Parkka et al. [369]	16 (3 F, 13 M)	21 activities in 7 locations; sitting at home, lying, sitting/reading newspaper, putting clothes on/going out, walking to bus stop, waiting for bus, traveling in bus, walking to restaurant, queuing, eating/drinking/talking, walking to library, sitting in library/reading, walking to shop, walking in shop/shopping, walking back home, Nordic walking, running, rowing machine, walking indoors, exercise bike, sitting/drinking	Wrist, chest (on rucksack strap), below left armpit (on breastbone), finger, forehead, shoulder (on rucksack strap), upper back (below neck)	22 wearables; air pressure, microphone, metal ball, 3-axial accelerometer (2 chest/wrist), 3D/2D compass (chest/wrist), voltage between EKG electrodes, humidity, light sensor, temperature sensor, switch, IR light absorption (heart rate/SaO2), IR light reflectance (heart rate/Pulse Plethysmogram/SaO2), voltage between chest belt electrodes, GPS satellite receiver, Piezo sensor, resistance between 2 metal leads, resistive temperature sensor	Automatically generated Decision Tree "overall classification accuracy" is 86%, custom Decision Tree is 82%, ANN is 82% Data recorded approximately 31 h age range 25.8 ± 4.3 years, body mass index (BMI) range 24.1 ± 3.0 kg/m^2
Ganti et al. [370]	2	7 stationary activities; sitting, reading, typing, lying down, standing in elevator, writing, eating with fork/knife, 3 non-stationary activities; walking, walking with umbrella, climbing stairs	Motes placed inside winter jacket; 2 motes placed on each arm (1 below/1 above elbow), 1 mote placed near waist	Five 2-axial accelerometer motes, GPS mote	HMM
Karantonis et al. [371]	6	12 activities; sit-to-stand, stand-to-sit, lying, lying-to-sit, sit-to-lying, walking (slow/normal/fast), fall (active/inactive/chair), circuit	Waist	3-axial accelerometer	Decision tree; overall accuracy 90.8%
Lyons et al. [372]	1	4 activities; sitting, standing, lying, moving	Trunk and Thigh	Two 2-axial ADXL202 accelerometers	Recording duration is 29 h and 28 min. Threshold based; detection accuracy; sitting 93%, standing 95%, lying 84%, moving 97%
Krause et al. [373], Study 1-Threshold analysis	11	Movement annotation (sitting, walking, running), Acoustic annotation (silence, conversation, music)	Armband at back of right upper arm (accelerometers), BlueSpoon headset (microphone)	Two 2-axial accelerometers, microphone	Threshold analysis
Ravi et al. [374]	2	8 activities; standing, walking, running, climbing up stairs, climbing down stairs, sitting-ups, vacuuming, brushing teeth	Near the pelvic region	3-axial CDXL04M3 accelerometer	Data of one subject in specific day used in training, while data of the other subject in another day used in testing. Tenfold cross-validation for every classifier. Naive Bayes (64%), SVM (63%), Decision tree (57%), k-NN (49.67%)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Krause et al. [373] , Study 2-Movement Study	1	6 types of movements; walking, running, sitting, Knee-bends, waving arms, climbing stairs	Armband at back of right upper arm (accelerometers)	Two 2-axial accelerometers	Data recorded over period of about 3 min. Kohonen Self-Organizing Map, k-Means clustering, Bayesian Networks
Mathie et al. [375]	26 (7 F, 19 M)	15 activities; motions/transitions activities (fall, walking, sit-to-stand, stand-to-sit, upright-to-lying, lying-to-upright, lying-to-lying, other movement) and postural orientations during rest (sitting, standing, lying back/front/left/right, inverted)	Waist	3-axial accelerometer	Age mean 30.5 ± standard deviation 6.3 years. Data set of 1309 movements. Hierarchical binary Decision Tree; overall sensitivity 97.7% and specificity 98.7%
Bao and Intille [60]	20 (7 F, 13 M)	20 activities; walking, walking/carry, sitting relaxed, computer work, standing still, eating/drinking, watching TV, reading, running, bicycling, stretching, strength train, scrubbing, vacuuming, folding laundry, lying down, brushing teeth, climbing stairs, riding elevator, riding escalator	Placed on the right hip, thigh, dominant wrist, non-dominant arm, ankle	5 ADXL210E 2-axial accelerometers	Age range 17–48 (mean 21.8, standard deviation 6.59) years. Each subject recorded 82–160 min. C4.5 Decision Tree classifier using leave-one-subject-out validation; overall accuracy 84.26%
Kern et al. [376]	1 (0 F, 1 M)	8 activities; sitting, standing, walking, stairs up, stairs down, shaking hands, writing on whiteboard, keyboard typing	All major joints on human body; just above ankle, just above knee, on hip, on wrist, just above elbow, on shoulder	12 three-axial accelerometers	Data recorded for 18.7 min. Bayesian classifier; overall accuracy is about 85%
Lee et al. [377]	24 (4 F, 20 M)	Static activities; standing up, sitting down with lowering subjects head, sitting down and leaning against, lying down straight, lying upside down. dynamic activities; running, going down, going up, walking	Back	3-axial accelerometer	Age range 21–28 years. adaptive threshold method; average classification rate 95.1%
Manty-jarvi et al. [5]	6	Walking in corridor, walking up/down stairs, opening doors	3 left hip, 3 right hip	6 3-axial ADXL202 accelerometers	Stratified tenfold cross-validation; 3 MLP neural networks; 83% to 90%
Van Laerhoven and Cakmakci [378]	1	7 activities; sitting, standing, walking, running, climbing stairs, descending stairs, riding bicycle	Outside of upper-leg, just above knee	Two 2-axial ADXL05s accelerometers	Kohonen Self-Organizing Map (SOM), Distance-weighted k-Nearest Neighbor (k-NN); recognition accuracy; 42% to 96%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Aminian et al. [379]	5 (1 F, 4 M)	4 activities; lying, sitting, standing, locomotion	Chest, thigh	2 one-axial accelerometers	Data recorded over a period of 1 h. Physical activity detection algorithm: 89.3%
Foerster et al. [380]	24 (0 F, 24 M)	9 activities: sitting, standing, lying, sitting/talking, sitting/operating PC, walking, stairs up, stairs down, cycling	Sternum, wrist, thigh, lower leg	4 accelerometers	Age range 21–34 years. Step-wise discriminatory analysis; overall accuracy 95.8%
Uiterwaal et al. [381]	1 (0 F, 1 M)	5 activities; lying, sitting, standing, locomotion, playing	2 on waist, 1 optional on left upper leg	Three 1-axis accelerometers	42 years old. Measurements for 3.5 h (maintenance mechanic work) and 3.2 h (local city hall messenger). ADL monitor; overall validity; 86% to 93%
Bouten et al. [382]	13 (13 Male)	dressing/breakfast, desk work, sitting, lying, making bed, stepping slow, standing/ walking, lunch, cleaning, leisure/snack, dinner, dish washing, stepping fast, carrying loads	Lower back	3-axial accelerometer	Age average 27 ± 4 years, body mass average 77 ± 12 kg, height average 1.83 ± 0.07 meter
Fall activities					
Mesanza et al. [383]	12 (4 F, 8 M)	User Fall dataset where subjects fall while using crutch; 8 static falls; standing still/take step/ trip over Assistive Devices for Walking (ADW)/fall forwards, fall backwards/simulating faint, rotate 90° to right/fall right, fall on right side, rotate 90° to left/fall left side, fall on left side. 8 dynamic falls; walk towards mattress/trip over ADW and fall forwards, walk towards mattress/simulate trip over object / fall forwards/left side/ right side/backwards, loss of balance/walking few meters/fall forwards/ backwards, walk/slide end up falling backwards. 7 regular activities using crutch; walking, walking faster, standing still, going up/down stairs, sit in chair, bend down to pick up object, balance loss without falling). ADW Fall dataset where crutch falls without user	GENEActiv; on chest, in pocket, on non-dominant wrist, on back. Sensorized Tip; on crutch	GENEActiv wearable sensors (four 3-axial accelerometers), Sensorized Tip (angle of inclination (α), 3-axial gyroscope, force sensor, 3-axial accelerometer)	Age range 25–40 years, 3 left-handed, 9 right-handed. dataset consists of 192 subject falls, 108 min physical activities, 5 min different static ADW positions/80 ADW falls. training dataset; 8 subjects, testing dataset; 4 subjects. SVM; Fall Detection F-score 0.963, precision 0.943, specificity 0.873, sensitivity 0.984

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Luna-Perejón et al [384]; AnkFALL	21 (5 F, 16 M)	11 activities; walking, sitting, getting up, crouching down, tripping over with left/right foot, walking while dizzy, falling backwards while sitting on void, tripping over and falling forward, falling without resisting to left/right	Ankle	One 3-axial ADXL345 accelerometer	Age range 21–60 years, height range 1.60–1.95 m, weight range 70–110 kg. Total dataset has 615 records. Hold-Out technique, Recurrent Neural Networks (GRU and LSTM); specificity; 92–93%
Chai et al. [385]	14 (14 M)	6 fall activities; forward fall with knees, forward fall with hands, left and right sides of inclined falls, backward fall, slow forward fall with crouch. 3 fall-like activities; crouching, sitting, walking with stoop	Chest IMU, 2 elbow IMUs, 2 wrist IMUs, 2 thigh IMUs, 2 ankle IMUs	9 IMUs	Age range 21–24 years, height range 1.7–1.88 m
Patel et al. [386]; VitaFALL		ADL activities; walking, stumbling, sitting. Falling activities; right, forward, backward, left	Wrist	3-axial ADXL335 accelerometer connected with Arduino Uno. Vital signs parameters; heart rate, heart beat, temperature monitoring	Directional falls detection accuracy 85%, specificity 100%, sensitivity 96%, Manual falls detection accuracy 85%, predictability 100%, sensitivity 100%
Palmerini et al. [291] (Falls)	40 (22 F, 18 M)	143 fall recordings	On the lower back	3-axial accelerometer	Most fall recordings are 40 min long. age range 69.2 ± 12.7 years. Subject-based cross-validation; fivefold cross-validation. Features from multiphase fall model + SVM; sensitivity 81.1%, false alarm rate per hour 0.56, F-measure 64.6%
Martínez-Villaseñor et al. [387]	17 (8 F, 9 M)	11 activities; 6 ADLs; walking, standing, sitting, picking up object, jumping, laying. 5 Falls; falling forward with hands, falling forward with knees, falling backwards, falling side-wards, falling sitting in empty chair	5 Mbientlab: left wrist, under neck, pants right pocket, waist middle in belt, left ankle. 1 electroencephalograph (EEG): at forehead. 6 infrared sensors: above room floor. 2 Microsoft cameras: at 1.82 m above floor	5 Mbientlab MetaSensor wearable sensors; 3-axis accelerometer, 3-axis gyroscope, ambient light value. 1 EEG NeuroSky MindWave headset. 6 infrared context-aware sensors. 2 Microsoft LifeCam Cinema cameras	Age range 18–24 years, height mean 1.66 m, weight mean 66.8 kg. Tenfold cross-validation; Best accuracy: IMU + EEG; 1 s window; Random Forest: 95.93 ± 0.30

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Alves et al. [297]; Dataset 1	19 (5 F, 14 M)	non-fall movements; seated/standing slowly/quickly drop sensor, stumble, get up chair, left/right lateral fall with recovery, walking/sitting on chair, lay/stand, forward/ backward fall with recovery, catch object floor, walk/run few meters, cough/sneeze, sit/stand, seated slightly lift body, laying on bed, bend/pick object floor, get up laying stand/sit, walk. fall movements; chair/stand up/ walk fall, backward fall ending lying/sitting, forward fall with rotation ending in lateral left/ right position, walk few meters then right/left lateral/ backward/forward fall, lateral fall to left/right ending lying flat, backward fall ending in lateral position, forward fall on knees/ending lying flat/forward arm protection	Chest, waist (fixed on belt), frontal pocket	Three 3-axial accelerometers	Age average 25 ± 2 years. 1399 non-fall movements (6.5 h) and 1009 simulated falls (4.5 h). state machine algorithm; fall detection rate 89.5%
Shahzad and Kim [388]; FallDroid	20 (3 F, 17 M)	19 activities; Fall events (forward improper weight shift, trip while walking forward, walking improper weight shift, trip while walking/turning, backward sitting, backward walking/slip, lateral left/right side, vertical collapse, fall freely). Fall-like ADLs (sitting, standing, lying down, rising up, pick up object from floor, pick something/continue walking/ stop, up/down stairs, jogging/ jumping, hitting sensor unintentionally)	Waist and thigh	Two LG G2 Smartphones; 3-axial accelerometer	Age range 28.45 ± 2.72, weight range 66.15 ± 10.83, height range 170.7 ± 7.68. waist: 209 falls and 137 fall-like ADLs. thigh: 175 falls and 196 fall-like ADLs. Threshold-based method (TBM) and multiple kernel learning SVM (MKL-SVM); fivefold cross-validation. waist and thigh: accuracy (97.81% and 91.70%), sensitivity (99.52% and 95.83%), specificity (95.19% and 88.01%)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Quadros et al. [389]	22	12 activities; falls (forward fall, backward fall, sideways fall (with device), sideways fall (without device), fall after rotating waist clockwise/counterclockwise) Non-Falls (walking, clapping hands, opening/closing door, moving object, tying shoes, sitting on chair)	Wrist	3-axial ADXL345 accelerometer, 3-axial L3G4200D gyroscope, 3-axial HMC5883L magnetometer	Age range 26.09 ± 4.73 years, height range 1.68 ± 0.11 m, weight range 67.82 ± 12.24 kg. Total number 792 signals; average signals duration 9.2 s. k-NN and fivefold cross-validation: accuracy 99.0%, sensitivity 100%, specificity 97.9%
Saha et al. [390]; University of Dhaka (DU)—Mobility Dataset (MD): DU-MD	25 (8 F, 17 M)	10 ADLs (7 basic ADLs and 3 falls); walking, sitting, lying, jogging, staircase climbing, staircase down, standing, falling (unconscious), falling (heart attack), falling (walk and fall)	Wrist-mounted	LIS2DH triaxial MEMS accelerometer	Age range 16–22 years, weight range 40–101 kg, height range 1.47–1.82 m
Chen et al. [391]	20 (5 F, 15 M)	6 ADLs: standing, sitting, standing up from chair, sitting down on chair, walking (0.8 m/s) 10 m, lying down on mattress	6 fall types: lateral falls to left/right followed by lying on floor, lateral falls to left/right followed by sitting-up on floor, falling forward and lying on floor, falling backward and lying on floor	Accelerometer	Age range 20–25 years, 12 trials. SVM; sensitivity 95%, specificity 96.7%
Casilari et al. [392]; UMAFall	17 (7 F, 10 M)	8 ADLs types: squatting, climbing stairs down/up, hopping, light jogging, lying down/getting up on/from a bed, sitting down/up on/from a chair, walking at normal pace. 3 Falls types: backwards, forwards, lateral	5 wearable sensing points; smartphone (right trouser pocket) and 4 wearable sensor motes (ankle, waist, right wrist, chest)	2 smartphone models; Samsung S5 and LG G4; 3-axis accelerometer. 4 wearable SensorTags motes; 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer	Age range 18–55 years, weight range 50–93 kg, height range 155–195 cm

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Sucerquia et al. [393]; SisFall	38 (19 F, 19 M)	19 ADLs; walking (slowly/quickly), jogging (slowly/quickly), walking upstairs/downstairs (slowly/quickly), sit half height chair/wait/get up (slowly/quickly), sit low height chair/wait/get up (slowly/quickly), sitting/trying get up/collapse, sitting/lying/wait/sit (slowly/quickly), one's back to lateral/wait/one's back, standing/slowly bending at/without knees/getting up, standing/into car/seated/out car, stumble while walking, jump without falling. 15 Falls; fall forward/backward slip, lateral fall slip, fall forward walking/jogging trip, vertical fall fainting, fall while walking, fall forward get up, lateral fall get up, fall forward/backward/lateral trying sit down, fall forward/backward/lateral sitting	Waist	2 accelerometers, gyroscope; only ADXL345 sensor acceleration data was used	23 adults age range 19–30 years, 15 elderly age range 60–75 years. adults height range 1.49–1.83 m, elderly height range 1.50–1.71 m. adults weight range 42–81 kg, elderly weight range 50–102 kg. Threshold-based classification, fall detection accuracy up to 96%
Klenk et al. [394]; FARSEEING	22 (12 F, 10 M)	22 falls; 21 falls indoor, 1 fall outdoor. activity before falling; 8 walking, 3 bending down, 4 transfer, 7 standing. falling direction; 3 forward, 7 backward, 2 sideways, 10 unknown	15 lumbar 5, 7 thigh	Accelerometer, gyroscope, magnetometer	Age mean: 69.5 years, height mean: 1.6713 m, weight mean: 75.5909 kg. FARSEEING (FAll Repository for the design of Smart and sElf-adaptive Environments prolonging Independent livinG)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Wang et al. [395]	11 (3 F, 8 M)	10 simulated falls; forward fall rotate 180° in air end facing up/down, forward fall during from sitting on chair to standing, forward fall during stepping down from higher platform, backward fall end facing up, backward fall attempting to sit on absent chair, backward fall after picking up something from floor, backward fall after turning around 180° in standing position, left/right lateral falls ending lying. 11 ADLs; bending down do laces then stand up again, sit on chair, jump vertically, pick up something from ground, walk/run for about 10 s, transitions from lie to sit to stand to walk, transition from standing to lying on ground, walk downstairs, wit on sofa, take elevator down four stories	Neck-worn lanyard	3-axial accelerometer, barometer	Height range 169.7 ± 8.1 cm, weight range 61.8 ± 10.8 kg, age range 25.9 ± 1.7 years. sensitivity 93.0% and specificity 87.3%
Vavoulas et al. [150]; MobiAct	57 (15 F, 42 M)	50 subjects completed successfully all 9 ADL types and 54 subjects completed all 4 fall types	For ADLs, phone is located in trousers pocket. For falls, phone is located in the pocket opposite to fall direction to protect phone from damage	Accelerometer, gyroscope, and orientation sensors of a Samsung Galaxy S3 smartphone with the LSM330DLC inertial module	Age range 20–47 years (average 26), height range 160–189 cm (average 175), weight range 50–120 kg (average 76)
Vilarinho et al. [396]; SINTEF ICT	3	12 falls; normal fall, step down from platform, self-tripping, falling backwards, falling backwards to sit down, falling backwards against wall, falling left, falling backwards/left, falling sideways left/right landing on wall, falling backwards/right, falling from sitting position. 7 ADLs; walking, turning around, sitting down slowly, sitting down quickly, tying shoes, stairs, jogging	2 sensors; 1 on thigh (in front pocket) and 1 on wrist	Android Wear smartwatch LG G Watch R and Android smartphone Samsung Galaxy S3; both devices have 3-axial accelerometer (utilized in experiments), 3-axial gyroscope, 3-axial compass	Age range 22–32, weight range 63–80 kg, height range 170–185 cm. Threshold-based and pattern recognition techniques; overall accuracy is 68%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Ojetola et al. [397]	42 (6 F, 36 M)	8 ADLs; standing, sitting on bed, sitting on chair, walking, crouching, lying, upstairs, downstairs. 4 Falls; forward, backward, lateral left, right	Chest, Thigh	2 Shimmer 3-axial accelerometers, 3-axial gyroscopes	Protocol 1; age 24.19 ± 6.19 years, height 172.66 ± 5.79 cm, weight 71.03 ± 11.86 kg, gender 4 F, 28 M. Protocol 2; age 23.9 ± 3.4 years, height 170.9 ± 8.8 cm, weight 65.5 ± 15.7 kg, gender 2 F, 8 M. Dataset 1; time per subject 23 min. Dataset 2; time per subject 4 min. J48 Decision Trees classifies falls and ADLs with F-measure 94.4%
Wertner et al. [398]	5	10 ADLs; walking, stopping, standing, sit down, sit down quickly, get up, stairs up, stairs down, bend forward, lay on floor. 4 falls; stumbling, slipping, sliding, get unconscious	On the hip in small bag	5 Different Android smartphone models; 3-axial accelerometer and 3-axial gyroscope	Recorded 6 h of falls and 4 h of ADLs
Shibuya et al. [399]; WGAS	1	Intentional falls, Activities of Daily Living (ADL), Dynamic Gait Index (DGI)	T4 or at waist level	3-axial accelerometer, 2 gyroscopes, MSP430 micro-controller	Support Vector Machine (SVM) classifier. 98.8% and 98.7% fall classification accuracies from data at T4 and belt positions. overall specificity 99.5% and overall sensitivity 97.0%
Fortino and Gravina [400]; Fall-Mobile-Guard	20	29 Activities; Falls patterns (front-lying, front-protecting-lying, front-knees-lying, front-right/left, front-quick-recovery, front-slow-recovery, back-sitting, back-lying, back-right/left, right/left-sideway, right/left-recovery, syncope, rolling-out-bed, falling-from-stairs), ADL patterns (lying-bed, rising-bed, sit-bed, sit-chair, sit-sofa, walking, walking-back, jogging, bending, bending-pick-up, trip-over, stumble, limp, squatting-down, coughing-sneezing)	Trousers pocket	One 3-axial Shimmer2R accelerometer	Age range 18–58 years, weight range 50–91 kg, height range 1.49–1.92 m. Accelerometer data of 340 falls and 320 ADL records. Threshold-based approach + k-NN; specificity 97%, sensitivity 83%, precision 90%
Pierleoni et al. [401]		Simulated falls, simulated falls with recovery, activities of daily living	Waist	3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer	
Sabatini et al. [402]; BIMU	25 (11 F, 14 M)	ADLs 7 types: walk-turn-walk, going up-and-down stairs, sweep, sit, stand-lie-stand, stand-sit-stand, pick-up. falls 5 types: forward fall, lateral fall, backward fall, forward trip, antero-lateral trip	The right anterior iliac spine	3-axis Accelerometer, 3-axis Gyroscope, barometric altimeter	Age range 25–35 years, body mass 50–116 kg, height 1.58–1.85 m. Single threshold Sensitivity = 80%, Specificity = 100%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Gjoreski et al. [403]; Evaluating Ambient Assisted Living—Activity Recognition (EvAAL-AR)	CMU &Utah: 15, JSI system: 10	7 activities: lying, sitting, standing, walking, bending, cycling on stationary bike, falling	CMU &Utah: pocket. JSI system: abdomen and thigh	Accelerometers, camera images, heart rate, breath rate, etc	CMU &Utah: the best recognition accuracy. JSI system: the highest final score
Kwolek and Kepski [404]	5	5 ADLs; walking up stairs, down stairs, picking up object, sitting down, standing up. 2 Falls; falling from standing, falling from sitting	Near pelvis	x-IMU inertial device (3-axial accelerometer, 3-axial gyroscope), Kinect sensor	Age over 26 years. SVM with tenfold cross-validation; fall detection accuracy 98.33%
Aziz et al. [405]	10	7 fall and 8 non-fall events	Waist	3-axial accelerometer, 3-axial gyroscope	Age range 22–32 years. Support vector machine (SVM) analysis. Sensitivity = 93.5 − 100%, Specificity = 85.6 − 99.7%
Medrano et al. [406]; tFall	10 (3 F, 7 M)	ADLs and 8 Falls; forward falls, backward falls, left/right lateral falls, syncope, sitting empty chair, forward without protection, falls with obstacle	2 pockets (left and right) and hand bag	Samsung Galaxy Mini phones; 3-axial accelerometer	Each fall is repeated 3 times per subject, and about 1 week for recording everyday behaviour. age range 20–42, 31.3 ± 8.6 years, body mass range 54–98, 69.2 ± 13.1 kg, height range 1.61 − 1.84, 1.73 ± 0.08 m. 1NN average performance; specificity 90% and sensitivity 91%, SVM achieved specificity 92% and sensitivity 95%
Ozdemir and Barshan [48]	14 (7 F, 7 M)	20 falls; front lying, front protecting lying, front knees, front knees lying, front right, front left, front quick recovery, front slow recovery, back sitting, back lying, back right, back left, right side way, right recovery, left side way, left recovery, syncope, syncope wall, podium, rolling out bed. 16 ADLs; lying bed, rising bed, sit bed, sit chair, sit sofa, sit air, walking forward, jogging, walking backward, bending, bending pick up, stumble, limp, squatting down, trip over, coughing sneezing	Head, chest, waist, right wrist, right thigh, right ankle	6 MTx wireless sensor units; each sensor unit 3-axial accelerometer, 3-axial gyroscope, 3-axial magnetometer and atmosphere pressure	Males (24 ± 3 years, 67.5 ± 13.5 kg, 172 ± 12 cm), females (21.5 ± 2.5 years, 58.5 ± 11.5 kg, 169.5 ± 12.5 cm). 1400 falls; 20 tasks × 14 subjects × 5 trials. 1120 ADLs; 16 tasks × 14 subjects × 5 trials. k-NN classifier and least squares method (LSM) achieve best sensitivity, specificity, accuracy all above 99%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Vavoulas et al. [407]; MobiFall	24 (7 F, 17 M)	9 participants performed falls and ADLs, while 15 performed only the different types of falls	For ADLs, phone is located in trousers' pocket. For falls, phone is located in the pocket opposite to fall direction	Accelerometer, gyroscope, and orientation sensors of a Samsung Galaxy S3 smartphone with the LSM330DLC inertial module	Males (age: 22–47 years, height: 1.69–1.89 m, weight: 62–103 kg) Females (age: 22–36 years, height: 1.60–1.72 m, weight: 50–90 kg)
Kepski et al. [408]; University of Rzeszow (UR) Fall Detection	30	Fall detection (lying on floor); dataset consisting of common ADLs, e.g., walking, taking/putting object from floor, bending right/left to lift object, sitting, tying laces, crouching down and lying	Microsoft Kinect cameras mounted at ceiling, 3-axial accelerometer worn near spine on lower back using elastic belt around waist	2 Microsoft Kinect cameras, 3-axial accelerometer	Age under 28 years. k-NN classifies lying pose versus common daily activities; tested on more than 45, 000 depth images with 0% false alarms
Tong et al. [409]	8	80 times of falls; 64 frontward falls, 16 side way falls. 40 times of each motion; standing (while talking), walking, sit-to-stand (on armchair), squat-to-stand, falling while walking frontward	Upper trunk; below the neck and above waist	3-axis Accelerometer	Age range 25 ± 3 years, weight range 61 ± 19 kg. Hidden Markov model (HMM); Sensitivity = 100% Specificity = 88.75%
Zhao et al. [410]	8	3 ADLs and 3 fall activities; forward, side way, backward fall	Chest	9 MTx IMU sensor modules	Age range 28.5 ± 4.3 years. Multiple threshold approach
Kerdegari et al. [411]	50 (32 F, 18 M)	20 activities; ADL patterns (walking, running, jumping, sitting chair, sitting ground, sitting-to-standing chair/ground, standing-to-lying, lying-to-standing), Fall patterns (forward/backward flexed knee, backward sitting on nothing, backward base on wall, side left/right flexed knee, side left/right base on wall, backward turning to left/right, fall freely)	Waist	One 3-axial ADXL345 accelerometer	Age average 32 years, height average 165 cm, weight average 66 kg. Acceleration data with total 6962 records and 29 features. Overall accuracy; Multilayer Perceptron 90.15%, Naive Bayes 88.57%, SVM 86.68%, C4.5 Decision Tree 84.28%, OneR 73.19%, ZeroR 66.49%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Kangas et al. [412]	16 (13 F, 3 M)	5 real life falls; subject entangled in blanket moving from bed to arm chair, subject entangled in blanket when moving from bed, subject found on floor near toilet, subject found lying on floor, subject likely fell out of bed lying on the floor next to bed	At waist in front of anterior superior iliac spine	3 axial accelerometer	Age average 88.4 ± 5.2 years
Lee and Carlisle [413]	18 (6 F, 12 M)	4 simulated falls; forwards, backwards, lateral left, lateral right. 8 ADLs; sit-to-stand, stand-to-sit, level walking, walking up/downstairs, answering the phone, picking up an object, getting up from supine	Waist	2 devices; mobile phone, independent accelerometer	Age range 29 ± 8.7 years. individual upper/lower detection thresholds; mobile phone: specificity 0.81 and sensitivity 0.77. external accelerometer: specificity 0.82 and sensitivity 0.96
Ojetola et al. [414]	8 (8 M)	4 types of falls; forward, backward, right, left	One sensor node on chest and the other on right thigh	2 SHIMMER sensor nodes; 3-axial accelerometers and 3-axial gyroscopes	Threshold-based methods. fall detection accuracy; over 90%, precision 81%, recall 92%
Shan and Yuan [415]	5 (5 M)	Falls; forward fall, backward fall, lateral fall left, lateral fall right. ADLs; sitting down, standing up, ascending stairs, descending stairs, object picking, walking, jogging, standing still	Waist	3-axial accelerometer	SVM; Sensitivity = 100%, Specificity = 100%
Curone et al. [416]	6	Upright standing, motionless lying down, upright mild activity, upright intense activity, active lying, standing to lying down, getting up from the ground, fall to the ground	Waist	3-axial accelerometer	Age range 22–37 years. Total recording 94 min; 336 steady actions (5373 samples) and 96 postural transitions (315 samples). threshold-based real-time self-calibrating algorithm; overall accuracy 96.2%
Dai et al. [417]	15 (2 F, 13 M)	Ffalls (forward, lateral, backward), ADLs (walking, jogging, standing, sitting)	Shirt pocket, on belt, or in pants pocket	Android G1 phone; 3-axial accelerometer	Age range 20–30 years. 3 subjects 161–170 cm, 7 subjects 171–180 cm, 5 subjects 181–190 cm. 1 subject < 50 kg, 2 subjects 51–60 kg, 5 subjects 61–70 kg, 7 subjects 71–80 kg. recorded data 450 falls; ADL data 20 min from each subject
Bianchi et al. [418]	20 (8 F, 12 M)	Real fall events and normal ADLs	Waist	3-axial MMA7260 accelerometer, barometric pressure SCP1000 sensor	Age range 23.7 ± 3.0 years. accuracy 96.9%, sensitivity 97.5%, specificity 96.5%; indoor environment
Yavuz et al. [419]	5	Falling, walking, jumping, sitting, lying on bed	Phone in pocket	Nexus One smartphone; accelerometer	Discrete wavelet transform (DWT)

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Li et al. [420]	3 (0 F, 3 M)	Fall patterns; e.g., falling forward/backward/leftward/ rightward/vertically/on stairs, etc. fall-like motions; e.g., sitting down fast, jumping, going up/downstairs, stumbling, lying down	Sensor nodes are attached on chest and thigh	Two TEMPO 3.0 sensor nodes, each includes; 3-axial accelerometer and 3-axial gyroscope	Subjects are in their 20 s. Fall ending with sitting sometimes triggers false negatives; sensitivity is 91%. Lying down quickly sometimes triggers false positives; specificity is 92%
Luštrek and Kaluža [421]	3	6 activities; falling, lying down, sitting down, standing/ walking, sitting and lying	12 body tags; attached to shoulders, elbows, wrists, hips, knees, ankles	12 radio tags	8 ML algorithms; C4.5 decision trees, RIPPER decision rules, Naive Bayes, 3-Nearest Neighbors, SVM, Random Forest, Bagging, AdaBoost M1 boosting. Tenfold cross-validation. SVM achieves the highest classification accuracy; on clean data 97.7%, on noisy data 96.5%
Kangas et al. [422]	41	6 simulated falls; syncope, tripping, sitting on empty air, slipping, lateral fall, rolling out of bed. 4 ADLs: sitting down on chair and getting up, picking up object from floor, lying down on bed and getting up, walking	Device attached with elastic belt at waist in front of anterior superior iliac spine	3-axial ADXL330 accelerometer	Simulated falls; 20 subjects; age range 40–65 years. ADLs; 21 subjects; age range 58–98 years. Falls recognition; sensitivity of 97.5% and specificity 100%
Sposaro and Tyson [423]	1	Falls and ADL activities; walking, running, standing	Front pant pocket	HTC G1 smartphone; 3-axial accelerometer	Threshold-based algorithms
Nyan et al. [424]	21	Falls and activities of daily living (ADL)	Torso and thigh	Two 3-axis Accelerometers, 3-axial Gyroscope, 2-axis Gyroscope	Multiple threshold approach, Sensitivity = 95.2% Specificity = 100%
Bourke et al. [425]	5	Falls and normal Activities of Daily Living (ADL)	Trunk and knees	3-axis accelerometer	Single threshold. Sensitivity = 100% Specificity = 100%
Kangas et al. [426]	3	Intentional falls; forward, backward, lateral	Waist, wrist, head	Three 3-axial accelerometers	Middle-aged subjects. threshold-based algorithms; sensitivity $97-98\%$ and specificity 100%
Wu and Xue [427]	24	Daily non-fall activities and simulated fall activities	On front of the waist	3-axis Accelerometer, 3-axis Gyroscope	10 young and 14 elderly subjects. Single threshold. Sensitivity = 100% Specificity = 100%
Bourke et al. [428]	10 (3 F, 7 M)	8 falls types; sitting down/standing up from armchair/kitchen chair/toilet seat/low stool, getting in/out of car seat, sitting down on/standing up from bed, lying down/standing up from bed, walking 10 ms	Trunk and thigh	Tri-axial accelerometer sensors	Age range 70–83 years. threshold-based fall detection algorithm
Jafari et al. [429]	2	68 falls types, sit-to-stand, stand-to-sit, lie-to-stand, stand-to-lie	The board is attached to the body	Wearable and Nokia 6680 smartphones; three-axis accelerometers	Age range 20–21 years. neural networks and k-Nearest Neighbors. average accuracy 84%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Kangas et al. [430]	2	Intentional falls (forward, backward, lateral) and ADLs	Waist, wrist, head	3-axial accelerometer	Specificity of fall detection; up to 100%
Linde-mann et al. [431]	2	Fall detector; ADLs and simulated falls	Head level; fixed behind the ear	Accelerometers	Young volunteer (intentional falls + ADLs). healthy elderly woman (83 years). Three trigger thresholds
Chen et al. [432]	2	Falling backwards, sitting, walking	Waist	2 orthogonal 2-axial ADXL210E accelerometers	Threshold based
Hand and finger gestures					
Zhuo et al. [433]	21 (6 F, 15 M)	Reading an article, scrolling social media, typing sentences, watching short videos, Non-use while sitting or walking	No restriction on how the subjects use the smartphone	Samsung Galaxy S7: IMU sensors (triaxial accelerometers, gyroscopes, and magnetometers)	Real-Time Smartphone Activity Classification. accuracy of 78.6% with the Extremely Randomized Trees algorithm; age range 21–42 years, mean age 25.1 and median age 23
Wen et al. [434]; Serendipity	10 (4 F, 6 M)	5 fine-grained gestures; pinch (to select), tap (click), rub fingers (scroll), squeeze (confirm), wave (decline)	Wrist	Samsung Galaxy Gear smartwatch; 3-axial accelerometer, 3-axial gyroscope, rotation, gravity sensors	Age range 20–37. SVM, Naive Bayes classifier, Logistic Regression, k-NN; SVM achieves best average f1-score of 87%
Georgi et al. [435]; Hand and Finger Gestures	5 (1 F, 4 M)	12 gestures; flick left, flick right, flick up, flick down, rotate left, rotate right, flat hand push, flat hand pull, palm pull, single click, double click, fist	IMU placed at the wrist on forearm top. 32 electrodes placed on forearm upper and lower sides	IMU records acceleration and angular velocity of forearm: 3D accelerometer and 3D gyroscope. EMG records electrical activity of forearm muscles during hand movements: 2 Plux biosignalsplux devices	Age range 23–34 years
Bulling et al. [436]	2	Human's arm movement; continuous sequence of gesture in daily living; 12 classes	6 body-worn sensors with sampling rate 32 HZ	Three 3-axial accelerometers, three 2-axial gyroscopes	Each sample has 15 dimensions. For each subject, all activities are repeated about 26 times
Bayati et al.[274]; HCI Gesture Dataset	1 (0 F, 1 M)	5 hand gestures; triangle, upside-down triangle, circle, square, infinity symbol	At various positions of right lower arm	6 USB acceleration sensors	For every action 50 repetitions are recorded. Expectation Maximization + adaptive LDA, average accuracy 86.9%
Junker et al. [437]	4 (1 F, 3 M)	Case study 1; shaking hands, light button, turn door knob/ open door, phone up/hold receiver to ear, open wallet with right hand/take coin/insert into vending machine. Case study 2; cutlery/fork and knife, pick-up cup/drink, soup using spoon, eat bread slice using hand only	Upper torso sensor, two upper arm sensors, two lower arm sensors	5 IMUs; Accelerometers and gyroscopes	Age range 25–35 years. relevant gestures; case study 1: 1826 s, case study 2: 5846 s. HMMs; classification accuracy: case study 1; 98.4%, case study 2; 97.4%

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Stiefmeier et al. [99, 103]; Skoda	1	10 gestures; write in notepad with right hand, opens/closes hood with left hand, gaps checking of front door with both hands, opens/closes car left front door, opens/closes 2 doors with both hands, gaps checking of trunk with both hands, opens trunk with both hands, grabs/turns steering wheel	Jacket with 20 sensor nodes cover the 2 arms and smartphone for data fusion and user interface	Sensor nodes are Zigbee MotionBee motes; 8 bit MCU, 3-axes accelerometer, Zigbee compliant radio	70 repetitions of each gesture. The dynamic classifier fusion core is "Hierarchical activity recognition algorithm", Node-level activity recognition is "Hidden Markov Models (HMM)", At network-level, node decisions are fused using "Naïve Bayesian approach"; recognition accuracy up to 97% using all available nodes
Healthcare applications, e.g., Parkinson's disease, rehabilitation activity, etc.					
Khowaja et al. [95]; mHealth dataset	10 (0 F, 10 M)	8 activities; standing, sitting, lying down, walking, walking upstairs, cycling, jogging, running	Chest, right wrist, left ankle	3 accelerometers, 2 gyroscopes, 2 magnetometers, 1 electrocardiogram (ECG) sensors	Error-correcting output coding (ECOC); Hierarchical classification model using QDF and Ensemble classifiers; overall accuracy 97.15%
Chen et al. [438]	5 (2 F, 3 M)	Medication adherence monitoring system for pill bottles: twist-cap and hand-to-mouth	IMU placed on wrist	Microsoft Kinect camera to segment the inertial sensor signal corresponding to twist-cap and hand-to-mouth by automatically time stamping the actions start and end. IMU (3-axis accelerometer, 3-axis gyroscope)	Dynamic Time Warping (DTW); twist-cap accuracy 95.0%, Hand-to-mouth accuracy 97.5%
Bachlin et al. [104]	10 Parkinson Disease patients (3 F, 7 M)	Wearable assistant for Parkinson's disease; 3 walking activities; walking back/forth in straight line, random walking in reception hall space, walking activities of ADL; entering/leaving rooms, walking to kitchen, getting drink, returning with cup of water	Shank (above ankle), thigh (above knee), at lower back attached to the same belt of wearable computer	3 tri-axial accelerometers	Age range 66.4 ± 4.8 years, Hoehn and Yahr score 2.6 ± 0.65. More than 8 h of recorded data. Detect Freezing of Gait (FOG) events using a "freeze" threshold; Freeze Index (FI) values above this threshold are identified as FOG events; "sensitivity" 73.1% and "specificity" 81.6%
Bachlin et al. [439]; Dynamic Analysis of Physiological Networks (DAPhNet) Gait dataset	10	Freezing of Gait (FOG); a common gait deficit in advanced Parkinson's disease (PD)	Shank sensor above the ankle, thigh sensor above the knee, trunk sensor and wearable computer attached to the belt, earphones around the neck	3-axis accelerometer, 2-axis gyroscope, 300mAh battery, Bluetooth radio	Subjects are idiopathic Parkinson's disease patients. Over 8 h of recorded data. Freeze of Gait detection; 73.1% sensitivity, 81.6% specificity

Table 15 (continued)

Dataset	Subjects	Activities	Sensor locations	Sensors	Comments
Jovanov et al. [440]		Computer assisted physical rehabilitation; remotely supervised rehabilitation activity/patient monitoring	Chest, wrist, ankles	Electrocardiogram (ECG) sensor to monitor heart activity, tilt sensor to monitor trunk position, SpO2 Oxygen saturation sensor, Motion sensors (accelerometers) to recognize human activity	

Acknowledgements This work is funded by the Information Technology Industry Development Agency (ITIDA), Information Technology Academia Collaboration (ITAC) Program, Egypt – Grant Number (PDP2023.R34.2 - VCOACH: Virtual Coaching for Indoor Exercising).

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data availability All data generated or analysed during this study are referenced in this published article (and its supplementary information files).

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ranasinghe S, Machot FA, Mayr HC (2016) A review on applications of activity recognition systems with regard to performance and evaluation. Int J Distrib Sens Netw 12(8):1550147716665520. https://doi.org/10.1177/1550147716665520
2. Ibrahim OT, Gomaa W, Youssef M (2019) CrossCount: a deep learning system for device-free human counting using WiFi. IEEE Sens J 19(21):9921–9928. https://doi.org/10.1109/JSEN.2019.2928502
3. Ibrahim OT, Gomaa W, Youssef M (2018) Zero-calibration device-free localization for the IoT based on participatory sensing. In: 2018 IEEE global communications conference (GLOBECOM), pp 1–7. https://doi.org/10.1109/GLOCOM.2018.8648111
4. Katz S, Chinn A, Cordrey L (1959) Multidisciplinary studies of illness in aged persons-ii: a new classification of functional status in activities of daily living. J Chronic Dis 9(1):55–62
5. Mantyjarvi J, Himberg J, Seppanen T (2001) Recognizing human motion with multiple acceleration sensors. In: IEEE international conference on systems, man, and cybernetics, vol 2. IEEE, pp 747–752
6. Turaga P, Chellappa R, Subrahmanian VS, Udrea O (2008) Machine recognition of human activities: a survey. IEEE Trans Circuits Syst Video Technol 18(11):1473–1488
7. Bruno B, Mastrogiovanni F, Sgorbissa A, Vernazza T, Zaccaria R (2012) Human motion modelling and recognition: a

computational approach. In: 2012 IEEE international conference on automation science and engineering (CASE). IEEE, pp 156–161
8. Bamberg SJM, Benbasat AY, Scarborough DM, Krebs DE, Paradiso JA (2008) Gait analysis using a shoe-integrated wireless sensor system. IEEE Trans Inf Technol Biomed 12(4):413–423
9. Maurer U, Rowe A, Smailagic A, Siewiorek DP (2006) ewatch: a wearable sensor and notification platform. In: International workshop on wearable and implantable body sensor networks. IEEE, p 4
10. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/
11. https://www.idtechex.com/fr/research-article/wearable-sensor-industry-worth-2-5bn-in-2020-new-idtechex-study-finds/21818
12. Zhang M, Sawchuk AA (2012) USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In: Proceedings of the 2012 ACM conference on ubiquitous computing. ACM, pp 1036–1043
13. Baños O, Damas M, Pomares H, Rojas I, Tóth MA, Amft O (2012) A benchmark dataset to evaluate sensor displacement in activity recognition. In: Proceedings of the 2012 ACM conference on ubiquitous computing. ACM, pp 1026–1035
14. Ashry S, Elbasiony R, Gomaa W (2018) An LSTM-based descriptor for human activities recognition using IMU sensors. In: Proceedings of the 15th international conference on informatics in control, automation and robotics ICINCO 2018, vol 1. SciTePress, INSTICC, pp 494–501. https://doi.org/10.5220/0006902404940501
15. Khan AM, Lee Y-K, Lee SY, Kim T-S (2010) A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer. IEEE Trans Inf Technol Biomed 14(5):1166–1172
16. Bruno B, Mastrogiovanni F, Sgorbissa A, Vernazza T, Zaccaria R (2013) Analysis of human behavior recognition algorithms based on acceleration data. In: 2013 IEEE international conference on robotics and automation (ICRA). IEEE, pp 1602–1607
17. Lara OD, Labrador MA (2012) A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor 15(3):1192–1209
18. Madcor H, Adel O, Gomaa W (2021) Location determination of on-body inertial sensors. In: Proceedings of the 18th international conference on informatics in control, automation and robotics—ICINCO. SciTePress, INSTICC, pp 693–700. https://doi.org/10.5220/0010582806930700
19. Mostafa A, Barghash TO, Assaf AA, Gomaa W (2020) Multi-sensor gait analysis for gender recognition. In: Proceedings of the 17th international conference on informatics in control, automation and robotics, vol 1. SciTePress, INSTICC, pp 629–636. https://doi.org/10.5220/0009792006290636
20. Adel O, Nafea Y, Hesham A, Gomaa W (2020) Gait-based person identification using multiple inertial sensors. In: Proceedings of the 17th international conference on informatics in control, automation and robotics, vol 1. SciTePress, INSTICC, pp 621–628. https://doi.org/10.5220/0009791506210628
21. Banos O, Toth MA, Damas M, Pomares H, Rojas I (2014) Dealing with the effects of sensor displacement in wearable activity recognition. Sensors 14(6):9995–10023
22. Chereshnev R, Kertész-Farkas A (2017) HuGaDB: human gait database for activity recognition from wearable inertial sensor networks. In: International conference on analysis of images, social networks and texts (AIST). Springer, pp 131–141
23. Permatasari J, Connie T, Song OT (2020) The MMUISD gait database and performance evaluation compared to public inertial sensor gait databases. In: Alfred R, Lim Y, Haviluddin H, On CK (eds) Computational science and technology. Springer, Singapore, pp 189–198
24. Gomaa W, Elbasiony R, Ashry S (2017) Adl classification based on autocorrelation function of inertial signals. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 833–837
25. Abdu-Aguye MG, Gomaa W, Makihara Y, Yagi Y (2019) On the feasibility of on-body roaming models in human activity recognition. In: Proceedings of the 16th international conference on informatics in control, automation and robotics, ICINCO 2019, pp 680–690. https://doi.org/10.5220/0007921606800690
26. Fayez A, Sharshar A, Hesham A, Eldifrawi I, Gomaa W (2022) Vais: a leading visual and inertial dataset of squats. In: 2022 International conference on ubiquitous information management and communication. IEEE, pp 1–8
27. Sharshar A, Fayez A, Eitta AA, Gomaa W (2022) MM-DOS: a novel dataset of workout activities. In: 2022 International joint conference on neural networks (IJCNN 2022), Padova, Italy. IEEE, pp 1–8
28. Margarito J, Helaoui R, Bianchi AM, Sartor F, Bonomi AG (2016) User-independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach. IEEE Trans Biomed Eng 63(4):788–796
29. Huynh T, Schiele B (2005) Analyzing features for activity recognition. In: Proceedings of the 2005 joint conference on smart objects and ambient intelligence: innovative context-aware services: usages and technologies. ACM, pp 159–163
30. Shumway RH, Stoffer DS (2017) Time series analysis and its applications: with R examples. Springer texts in statistics, Paperback edn. Springer, p 575
31. Strömbäck D, Huang S, Radu V (2020) MM-Fit: multimodal deep learning for automatic exercise logging across sensing devices. Proc ACM InteractMob Wearable Ubiquitous Technol 4(4):1–22
32. Mostafa A, Elsagheer S, Gomaa W (2021) BioDeep: a deep learning system for IMU-based human biometrics recognition. In: Proceedings of the 18th international conference on informatics in control, automation and robotics—ICINCO. SciTePress, INSTICC, pp 620–629. https://doi.org/10.5220/0010578806200629
33. Ngo TT, Atiqur Rahman Ahad M, Antar AD, Ahmed M, Muramatsu D, Makihara Y, Yagi Y, Inoue S, Hossain T, Hattori Y (2019) OU-ISIR wearable sensor-based gait challenge: age and gender. In: 2019 International conference on biometrics (ICB), pp 1–6. https://doi.org/10.1109/ICB45273.2019.8987235
34. Miraldo DC, Watanabe RN, Duarte M (2020) An open data set of inertial, magnetic, foot-ground contact, and electromyographic signals from wearable sensors during walking. Motor Control 24:1–13
35. Nourani H (2020) A comprehensive comparison of human activity recognition using inertial sensors. PhD thesis, Concordia University
36. Wang Y (2019) A data fusion-based hybrid sensory system for older people's daily activity recognition. PhD thesis, Bournemouth University
37. Janidarmian M, Roshan Fekr A, Radecka K, Zilic Z (2017) A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors 17(3):529
38. Morris D, Saponas TS, Guillory A, Kelner I (2014) Recofit: using a wearable sensor to find, recognize, and count repetitive exercises. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, pp 3225–3234
39. Muehlbauer M, Bahle G, Lukowicz P (2011) What can an arm holster worn smart phone do for activity recognition? In: 2011 15th Annual international symposium on wearable computers. IEEE, pp 79–82

40. Zebin T, Scully PJ, Ozanyan KB (2017) Inertial sensor based modelling of human activity classes: feature extraction and multi-sensor data fusion using machine learning algorithms. In: eHealth 360°. Springer, pp 306–314
41. Jain R, Semwal VB, Kaushik P (2021) Stride segmentation of inertial sensor data using statistical methods for different walking activities. Robotica 40:1–14
42. Reyes-Ortiz J-L, Oneto L, Samà A, Parra X, Anguita D (2016) Transition-aware human activity recognition using smartphones. Neurocomputing 171:754–767
43. Ngo TT, Makihara Y, Nagahara H, Mukaigawa Y, Yagi Y (2014) The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. Pattern Recognit 47(1):228–237. https://doi.org/10.1016/j.patcog.2013.06.028
44. Bota P, Silva J, Folgado D, Gamboa H (2019) A semi-automatic annotation approach for human activity recognition. Sensors 19(3):501
45. Anguita D, Ghio A, Oneto L, Parra Perez X, Reyes Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In: Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning (ESANN), pp 437–442
46. Altun K, Barshan B (2010) Human activity recognition using inertial/magnetic sensor units. In: International workshop on human behavior understanding. Springer, pp 38–51
47. Yu X (2019) Human activity recognition using wearable inertia sensor data and machine learning. PhD thesis, Purdue University Graduate School
48. Özdemir AT, Barshan B (2014) Detecting falls with wearable sensors using machine learning techniques. Sensors 14(6):10691–10708
49. Pereira A, Folgado D, Cotrim R, Sousa I (2019) Physiotherapy exercises evaluation using a combined approach based on sEMG and wearable inertial sensors. In: Proceedings of the 12th international joint conference on biomedical engineering systems and technologies—biosignals. SciTePress, pp 73–82
50. Rosati S, Balestra G, Knaflitz M (2018) Comparison of different sets of features for human activity recognition by wearable sensors. Sensors 18(12):4189
51. Muaaz M, Chelli A, Abdelgawwad AA, Mallofré AC, Pätzold M (2020) Wiwehar: multimodal human activity recognition using Wi-Fi and wearable sensing modalities. IEEE Access 8:164453–164470
52. Abdu-Aguye MG, Gomaa W (2018) Novel approaches to activity recognition based on vector autoregression and wavelet transforms. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 951–954
53. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
54. Rainforth T, Wood FD (2015) Canonical correlation forests. arXiv:abs/1507.05444
55. Gupta C, Suggala AS, Goyal A, Simhadri HV, Paranjape B, Kumar A, Goyal S, Udupa R, Varma M, Jain P (2017) ProtoNN: compressed and accurate kNN for resource-scarce devices. In: Precup D, Teh YW (eds) Proceedings of the 34th international conference on machine learning. Proceedings of machine learning research, vol 70. PMLR, pp 1331–1340. https://proceedings.mlr.press/v70/gupta17a.html
56. Zhou Z-H, Feng J (2017) Deep forest: towards an alternative to deep neural networks. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI-17, pp 3553–3559. https://doi.org/10.24963/ijcai.2017/497
57. Altun K, Barshan B, Tunçel O (2010) Comparative study on classifying human activities with miniature inertial and magnetic sensors. Pattern Recognit 43(10):3605–3620
58. Preece SJ, Goulermas JY, Kenney LP, Howard D (2008) A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans Biomed Eng 56(3):871–879
59. Foerster F, Fahrenberg J (2000) Motion pattern and posture: correctly assessed by calibrated accelerometers. Behav Res Methods Instrum Comput 32(3):450–457
60. Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: International conference on pervasive computing. Springer, pp 1–17
61. Eyobu OS, Han DS (2018) Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. Sensors 18(9):2892
62. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2012) Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: International workshop on ambient assisted living. Springer, pp 216–223
63. Madrid García A (2016) Human activity recognition by inertial signals obtained from a smartphone
64. Machado IP, Gomes AL, Gamboa H, Paixão V, Costa RM (2015) Human activity data discovery from triaxial accelerometer sensor: non-supervised learning sensitivity to feature extraction parametrization. Inf Process Manag 51(2):204–214
65. Krause A, Siewiorek DP, Smailagic A, Farringdon J (2003) Unsupervised, dynamic identification of physiological and activity context in wearable computing. In: Seventh IEEE international symposium on wearable computers (ISWC). IEEE, pp 88–97
66. Abreu M, Barandas M, Leonardo R, Gamboa H (2019) Detailed human activity recognition based on multiple hmm. In: Proceedings of the 12th international joint conference on biomedical engineering systems and technologies—BIOSIGNALS. SciTePress, pp 171–178
67. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors 14(6):10146–10176
68. Tahir SBUD, Jalal A, Kim K (2020) Wearable inertial sensors for daily activity analysis based on Adam optimization and the maximum entropy Markov model. Entropy 22(5):579
69. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980
70. Osterland S, Weber J (2019) Analytical analysis of single-stage pressure relief valves. Int J Hydromechatron 2(1):32–53
71. Jalal A, Quaid MAK, Hasan AS (2018) Wearable sensor-based human behavior understanding and recognition in daily life for smart environments. In: 2018 International conference on frontiers of information technology (FIT). IEEE, pp 105–110
72. Banos O, Villalonga C, Garcia R, Saez A, Damas M, Holgado-Terriza JA, Lee S, Pomares H, Rojas I (2015) Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed Eng Online 14(2):1–20
73. Figueira C, Matias R, Gamboa H (2016) Body location independent activity monitoring. In: Proceedings of the 9th international joint conference on biomedical engineering systems and technologies—BIOSIGNALS, vol 5. SciTePress, pp 190–197
74. Attal F, Mohammed S, Dedabrishvili M, Chamroukhi F, Oukhellou L, Amirat Y (2015) Physical human activity recognition using wearable sensors. Sensors 15(12):31314–31338
75. Suto J, Oniga S, Sitar PP (2016) Feature analysis to human activity recognition. Int J Comput Commun Control 12(1):116–130

76. Yang AY, Jafari R, Sastry SS, Bajcsy R (2009) Distributed recognition of human actions using wearable motion sensor networks. J Ambient Intell Smart Environ 1(2):1–5
77. He Z (2010) Activity recognition from accelerometer signals based on wavelet-AR model. In: 2010 IEEE international conference on progress in informatics and computing, vol 1, pp 499–502. https://doi.org/10.1109/PIC.2010.5687572
78. Hassan MM, Uddin MZ, Mohamed A, Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. Future Gener Comput Syst 81:307–313
79. San-Segundo R, Montero JM, Barra-Chicote R, Fernández F, Pardo JM (2016) Feature extraction from smartphone inertial signals for human activity segmentation. Signal Process 120:359–372
80. Hermansky H, Morgan N (1994) Rasta processing of speech. IEEE Trans Speech Audio Process 2(4):578–589
81. Anguita D, Ghio A, Oneto L, Llanas Parra FX, Reyes Ortiz JL (2013) Energy efficient smartphone-based activity recognition using fixed-point arithmetic. J Univ Comput Sci 19(9):1295–1314
82. Reyes-Ortiz JL, Ghio A, Parra X, Anguita D, Cabestany J, Catala A (2013) Human activity and motion disorder recognition: towards smarter interactive cognitive environments. In: 21th European symposium on artificial neural networks, computational intelligence and machine learning, ESANN. Citeseer
83. Ashry S, Gomaa W, Abdu-Aguye MG, El-borae N (2020) Improved IMU-based human activity recognition using hierarchical hmm dissimilarity. In: Proceedings of the 17th international conference on informatics in control, automation and robotics, vol 1, ICINCO. SciTePress, INSTICC, pp 702–709. https://doi.org/10.5220/0009886607020709
84. Gomaa W (2019) Statistical and time series analysis of accelerometer signals for human activity recognition. In: 2019 14th International conference on computer engineering and systems (ICCES), pp 351–356
85. Schay G (2007) Introduction to probability with statistical applications, 2007 edn. Birkhäuser. http://amazon.com/o/ASIN/0817644970/
86. Gibbons JD, Chakraborti S (2010) Nonparametric statistical inference, 15th edn. Statistics: textbooks and monographs. Chapman and Hall/CRC. http://amazon.com/o/ASIN/1420077619/
87. Bruno B, Mastrogiovanni F, Sgorbissa A, Vernazza T, Zaccaria R (2013) Analysis of human behavior recognition algorithms based on acceleration data. In: 2013 IEEE international conference on robotics and automation. IEEE, pp 1602–1607. https://doi.org/10.1109/ICRA.2013.6630784
88. Gomaa W (2020) Statistical metric-theoretic approach to activity recognition based on accelerometer data. In: Hassanien AE, Shaalan K, Tolba MF (eds) Proceedings of the international conference on advanced intelligent systems and informatics 2019. Springer, Cham, pp 537–546
89. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. science 313(5786):504–507
90. Abdu-Aguye MG, Gomaa W (2019) VersaTL: versatile transfer learning for IMU-based activity recognition using convolutional neural networks. In: Proceedings of the 16th international conference on informatics in control, automation and robotics, ICINCO 2019, vol 1, Prague, Czech Republic, July 29–31, 2019, pp 507–516. https://doi.org/10.5220/0007916705070516
91. Reyes-Ortiz J-L, Oneto L, Ghio A, Samá A, Anguita D, Parra X (2014) Human activity recognition on smartphones with awareness of basic activities and postural transitions. In: International conference on artificial neural networks. Springer, pp 177–184
92. Abdu-Aguye MG, Gomaa W (2019) Robust human activity recognition based on deep metric learning. In: Proceedings of the 16th international conference on informatics in control, automation and robotics, ICINCO 2019, vol 1, Prague, Czech Republic, July 29–31, 2019, pp 656–663. https://doi.org/10.5220/0007916806560663
93. Khaertdinov B, Ghaleb E, Asteriadis S (2021) Deep triplet networks with attention for sensor-based human activity recognition. In: 2021 IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 1–10. https://doi.org/10.1109/PERCOM50583.2021.9439116
94. Reiss A, Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In: Proceedings of the 2012 16th international symposium on wearable computers. IEEE, pp 108–109
95. Khowaja SA, Yahya BN, Lee S-L (2017) Hierarchical classification method based on selective learning of slacked hierarchy for activity recognition systems. Expert Syst Appl 88:165–177
96. Martinez J, Hossain R, Romero J, Little J (2017) A simple yet effective baseline for 3d human pose estimation. In; 2017 IEEE international conference on computer vision (ICCV). IEEE, pp 2659–2668
97. Mahmud S, Tonmoy MTH, Bhaumik KK, Rahman AM, Amin MA, Shoyaib M, Khan MAH, Ali AA (2020) Human activity recognition from wearable sensor data using self-attention. In: ECAI 2020. IOS Press, pp 1332–1339
98. Roggen D, Calatroni A, Rossi M, Holleczek T, Förster K, Tröster G, Lukowicz P, Bannach D, Pirkl G, Ferscha A et al (2010) Collecting complex activity datasets in highly rich networked sensor environments. In: 2010 Seventh international conference on networked sensing systems (INSS). IEEE, pp 233–240
99. Stiefmeier T, Roggen D, Ogris G, Lukowicz P, Tröster G (2008) Wearable activity tracking in car manufacturing. IEEE Pervasive Comput 7(2):42–50
100. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems, vol 27
101. Tao W, Chen H, Moniruzzaman M, Leu MC, Yi Z, Qin R (2021) Attention-based sensor fusion for human activity recognition using IMU signals. arXiv preprint arXiv:2112.11224
102. Barshan B, Yüksek MC (2014) Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. Comput J 57(11):1649–1667
103. Zappi P, Roggen D, Farella E, Tröster G, Benini L (2012) Network-level power-performance trade-off in wearable activity recognition: a dynamic sensor selection approach. ACM Trans Embed Comput Syst (TECS) 11(3):1–30
104. Bachlin M, Plotnik M, Roggen D, Maidan I, Hausdorff JM, Giladi N, Troster G (2010) Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. IEEE Trans Inf Technol Biomed 14(2):436–446
105. Li Y, Wang L (2022) Human activity recognition based on residual network and BiLSTM. Sensors 22(2):635
106. Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. ACM SIGKDD Explor Newsl 12(2):74–82
107. Mondal R, Mukherjee D, Singh PK, Bhateja V, Sarkar R (2020) A new framework for smartphone sensor-based human activity recognition using graph neural network. IEEE Sens J 21(10):11461–11468
108. Raza A, Tran KP, Koehl L, Li S, Zeng X, Benzaidi K (2021) Lightweight transformer in federated setting for human activity recognition. arXiv preprint arXiv:2110.00244

109. Sozinov K, Vlassov V, Girdzijauskas S (2018) Human activity recognition using federated learning. In: 2018 IEEE international conference on parallel and distributed processing with applications, ubiquitous computing and communications, big data and cloud computing, social computing and networking, sustainable computing and communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom). IEEE, pp 1103–1111
110. Ignatov A (2018) Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl Soft Comput 62:915–922. https://doi.org/10.1016/j.asoc.2017.09.027
111. Murad A, Pyun J-Y (2017) Deep recurrent neural networks for human activity recognition. Sensors 17(11):2556
112. EK S, Portet F, Lalanda P (2022) Lightweight transformers for human activity recognition on mobile devices. arXiv preprint arXiv:2209.11750
113. Mutegeki R, Han DS (2020) A CNN-LSTM approach to human activity recognition. In: 2020 International conference on artificial intelligence in information and communication (ICAIIC). IEEE, pp 362–366
114. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S et al (2021) An image is worth 16 × 16 words: transformers for image recognition at scale. In: International conference on learning representations
115. Mehta S, Rastegari M (2022) MobileVit: light-weight, general-purpose, and mobile-friendly vision transformer. In: International conference on learning representations
116. Augustinov G, Nisar MA, Li F, Tabatabaei A, Grzegorzek M, Sohrabi K, Fudickar S (2022) Transformer-based recognition of activities of daily living from wearable sensor data. In: Proceedings of the 7th international workshop on sensor-based activity recognition and artificial intelligence, pp 1–8
117. Li B, Cui W, Wang W, Zhang L, Chen Z, Wu M (2021) Two-stream convolution augmented transformer for human activity recognition. In: Proceedings of the AAAI conference on artificial intelligence, vol 35, pp 286–293
118. Yousefi S, Narui H, Dayal S, Ermon S, Valaee S (2017) A survey on behavior recognition using WiFi channel state information. IEEE Commun Mag 55(10):98–104
119. Wang W, Liu AX, Shahzad M, Ling K, Lu S (2017) Device-free human activity recognition using commercial WiFi devices. IEEE J Sel Areas Commun 35(5):1118–1131
120. Chen Z, Zhang L, Jiang C, Cao Z, Cui W (2018) WiFi CSI based passive human activity recognition using attention based BLSTM. IEEE Trans Mob Comput 18(11):2714–2724
121. Dirgová Luptáková I, Kubovčík M, Pospíchal J (2022) Wearable sensor-based human activity recognition with transformer model. Sensors 22(5):1911
122. Wang H, Zhao J, Li J, Tian L, Tu P, Cao T, An Y, Wang K, Li S (2020) Wearable sensor-based human activity recognition using hybrid deep learning techniques. Secur Commun Netw 2020:1–12
123. Sikder N, Nahid A-A (2021) KU-HAR: an open dataset for heterogeneous human activity recognition. Pattern Recognit Lett 146:46–54
124. Zhu W, Lan C, Xing J, Zeng W, Li Y, Shen L, Xie X (2016) Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. In: Thirtieth AAAI conference on artificial intelligence
125. Soomro K, Idrees H, Shah M (2019) Online localization and prediction of actions and interactions. IEEE Trans Pattern Anal Mach Intell 41(2):459–472
126. Aruba: Tulum Datasets from WSU CASAS Smart home project. http://ailab.wsu.edu/casas/datasets/
127. Cook D, Schmitter-Edgecombe M, Crandall A, Sanders C, Thomas B (2009) Collecting and disseminating smart home sensor data in the casas project. In: Proceedings of the CHI workshop on developing shared home behavior datasets to advance HCI and ubiquitous computing research. ACM, pp 1–7
128. Yala N, Fergani B, Fleury A (2017) Towards improving feature extraction and classification for activity recognition on streaming data. J Ambient Intell Humaniz Comput 8(2):177–189
129. Ashry S, Ogawa T, Gomaa W (2020) CHARM-deep: Continuous human activity recognition model based on deep neural network using IMU sensors of smartwatch. IEEE Sens J 20(15):8757–8770
130. Khannouz M, Glatard T (2020) A benchmark of data stream classification for human activity recognition on connected objects. Sensors 20(22):6486
131. Krishnan NC, Cook DJ (2014) Activity recognition on streaming sensor data. Pervasive Mob Comput 10:138–154
132. Abdallah ZS, Gaber MM, Srinivasan B, Krishnaswamy S (2015) Adaptive mobile activity recognition system with evolving data streams. Neurocomputing 150:304–317
133. Do TM, Loke SW, Liu F (2012) Healthylife: an activity recognition system with smartphone using logic-based stream reasoning. In: International conference on mobile and ubiquitous systems: computing, networking, and services. Springer, pp 188–199
134. Abdallah ZS, Gaber MM, Srinivasan B, Krishnaswamy S (2016) Anynovel: detection of novel concepts in evolving data streams. Evol Syst 7(2):73–93
135. Roggen D, Forster K, Calatroni A, Holleczek T, Fang Y, Troster G, Ferscha A, Holzmann C, Riener A, Lukowicz P et al (2009) OPPORTUNITY: towards opportunistic activity and context recognition systems. In: 2009 IEEE international symposium on a world of wireless, mobile and multimedia networks and workshops. IEEE, pp 1–6
136. Yala N, Fergani B, Fleury A (2015) Feature extraction for human activity recognition on streaming data. In: 2015 International symposium on innovations in intelligent systems and applications (INISTA). IEEE, pp 1–6
137. Zhang Y, Ramachandran KM (2019) Human activity recognition with streaming smartphone data. In: 2019 Global conference for advancement in technology (GCAT). IEEE, pp 1–6
138. Kwon H, Tong C, Haresamudram H, Gao Y, Abowd GD, Lane ND, Ploetz T (2020) Imutube: automatic extraction of virtual on-body accelerometry from video for human activity recognition. Proc ACM Interact Mob Wearable Ubiquitous Technol 4(3):1–29
139. Goodfellow I, Bengio Y, Courville A (2016) Deep learning, Illustrated. Adaptive computation and machine learning series. The MIT Press, Cambridge
140. Olivas ES, Guerrero JDM, Martinez Sober M, Magdalena Benedito JR, Lopez AJS (2009) Handbook of research on machine learning applications and trends—algorithms, methods, and techniques, Hershey, PA. Information science reference
141. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22(10):1345–1359. https://doi.org/10.1109/TKDE.2009.191
142. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ (eds) Advances in neural information processing systems, vol 27. Curran Associates, Inc., pp 3320–3328. https://proceedings.neurips.cc/paper/2014/file/375c71349b295fbe2dcdca9206f20a06-Paper.pdf
143. Hu DH, Zheng VW, Yang Q (2011) Cross-domain activity recognition via transfer learning. Pervasive Mob Comput 7(3):344–358. https://doi.org/10.1016/j.pmcj.2010.11.005

144. Khan MAAH, Roy N (2017) TransAct: transfer learning enabled activity recognition. In: 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops). IEEE, pp 545–550. https://doi.org/10.1109/PERCOMW.2017.7917621
145. Khan MAAH, Roy N (2018) 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). IEEE, pp 37–47. https://doi.org/10.1109/IoTDI.2018.00014
146. Chavarriaga R, Sagha H, Calatroni A, Digumarti ST, Tröster G, Millán JdR, Roggen D (2013) The opportunity challenge: a benchmark database for on-body sensor-based activity recognition. Pattern Recognit Lett 34(15):2033–2042
147. Pan SJ, Tsang IW, Kwok JT, Yang Q (2010) Domain adaptation via transfer component analysis. IEEE Trans Neural Netw 22(2):199–210
148. Long M, Wang J, Ding G, Sun J, Yu PS (2013) Transfer feature learning with joint distribution adaptation. In: Proceedings of the IEEE international conference on computer vision. IEEE, pp 2200–2207
149. Chikhaoui B, Gouineau F, Sotir M (2018) A CNN based transfer learning model for automatic activity recognition from accelerometer sensors. In: Machine learning and data mining in pattern recognition: 14th international conference, MLDM 2018, New York, NY, USA, July 15–19, 2018, proceedings, Part II 14. Springer, pp 302–315
150. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M (2016) The MobiAct dataset: recognition of activities of daily living using smartphones. In: International conference on information and communication technologies for ageing well and e-health, vol 2. SCITEPRESS, pp 143–151
151. Sztyler T, Stuckenschmidt H, Petrich W (2017) Position-aware activity recognition with wearable devices. Pervasive Mob Comput 38:281–295
152. Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, Sonne T, Jensen MM (2015) 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. ACM, pp 127–140
153. Oh C, Hwang H, Lee H-y, Lim Y, Jung G, Jung J, Choi H, Song K (2023) Blackvip: black-box visual prompting for robust transfer learning. arXiv preprint arXiv:2303.14773
154. Sharshar A, Eitta AA, Fayez A, Khaims MA, Zaki AB, Gomaa W (2023) Camera coach: activity recognition and assessment using thermal and RGB videos. In: International joint conference on neural networks (IJCNN 2023), Queensland, Australia. IEEE
155. Zou X, Yin D, Zhong Q, Yang H, Yang Z, Tang J (2021) Controllable generation from pre-trained language models via inverse prompting. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery and data mining. ACM, pp 2450–2460
156. Li J, Tang T, Nie J-Y, Wen J-R, Zhao, WX (2022) Learning to transfer prompts for text generation. arXiv preprint arXiv:2205.01543
157. Wang Z, Panda R, Karlinsky L, Feris R, Sun H, Kim Y (2023) Multitask prompt tuning enables parameter-efficient transfer learning. In: International conference on learning representations (ICLR 2023), Kigali, Rwanda
158. Markitantov M (2020) Transfer learning in speaker's age and gender recognition. In: Karpov A, Potapova R (eds) Speech and computer. Springer, Cham, pp 326–335
159. Albuquerque L, Oliveira C, Teixeira A, Figueiredo D (2021) Eppur si muove: formant dynamics is relevant for the study of speech aging effects. In: Biosignals
160. Sun M, Li C, Zha H (2017) Inferring private demographics of new users in recommender systems. In: Proceedings of the 20th ACM international conference on modelling, analysis and simulation of wireless and mobile systems. MSWiM '17, New York, NY, USA. Association for Computing Machinery, pp 237–244. https://doi.org/10.1145/3127540.3127566
161. Rosli N, Rahman M, Balakrishnan M, Komeda T, Mazlan S, Zamzuri H (2017) Improved gender recognition during stepping activity for rehab application using the combinatorial fusion approach of EMG and HRV. Appl Sci. https://doi.org/10.3390/app7040348
162. Van Hamme T, Garofalo G, Argones Rúa E, Preuveneers D, Joosen W (2019) A systematic comparison of age and gender prediction on IMU sensor-based gait traces. Sensors. https://doi.org/10.3390/s19132945
163. Riaz Q, Vögele A, Krüger B, Weber A (2015) One small step for a man: estimation of gender, age and height from recordings of one step by a single inertial sensor. Sensors 15(12):31999–32019. https://doi.org/10.3390/s151229907
164. Jain A, Kanhangad V (2016) Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings. In: 2016 International conference on computational techniques in information and communication technologies (ICCTICT). IEEE, pp 597–602. https://doi.org/10.1109/ICCTICT.2016.7514649
165. Bengio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127. https://doi.org/10.1561/2200000006
166. Riaz Q, Hashmi MZUH, Hashmi MA, Shahzad M, Errami H, Weber A (2019) Move your body: age estimation based on chest movement during normal walk. IEEE Access 7:28510–28524
167. Khabir KM, Siraj MS, Ahmed M, Ahmed MU (2019) Prediction of gender and age from inertial sensor-based gait dataset. In: 2019 Joint 8th international conference on informatics, electronics and vision (ICIEV) and 2019 3rd international conference on imaging, vision and pattern recognition (icIVPR). IEEE, pp 371–376
168. Rasnayaka S, Sim T (2020) Your tattletale gait privacy invasiveness of IMU gait data. In: 2020 IEEE international joint conference on biometrics (IJCB). IEEE, pp 1–10
169. Murray MP, Drought AB, Kory RC (1964) Walking patterns of normal men. JBJS 46(2):335–360
170. Singh JP, Jain S, Arora S, Singh UP (2018) Vision-based gait recognition: a survey. IEEE Access 6:70497–70527. https://doi.org/10.1109/ACCESS.2018.2879896
171. Sprager S, Juric MB (2015) Inertial sensor-based gait recognition: a review. Sensors (Basel, Switzerland) 15:22089–22127
172. Cola G, Avvenuti M, Musso F, Vecchio A (2016) Gait-based authentication using a wrist-worn device. In: Proceedings of the 13th international conference on mobile and ubiquitous systems: computing, networking and services. ACM, pp 208–217
173. Zhao Y, Zhou S (2017) Wearable device-based gait recognition using angle embedded gait dynamic images and a convolutional neural network. Sensors 17(3):478
174. Derawi MO, Nickel C, Bours P, Busch C (2010) Unobtrusive user-authentication on mobile phones using biometric gait recognition. In: 2010 Sixth international conference on intelligent information hiding and multimedia signal processing. IEEE, pp 306–311
175. Shen C, Chen Y, Guan X (2018) Performance evaluation of implicit smartphones authentication via sensor-behavior analysis. Inf Sci 430:538–553

176. Adel O, Soliman M, Gomaa W (2021) Inertial gait-based person authentication using Siamese networks. In: 2021 International joint conference on neural networks (IJCNN), pp 1–7. https://doi.org/10.1109/IJCNN52387.2021.9534261
177. Shield A, Zhou S (2004) Assessing voluntary muscle activation with the twitch interpolation technique. Sports Med 34(4):253–267. https://doi.org/10.2165/00007256-200434040-00005
178. Lucas KR, Rich PA, Polus BI (2010) Muscle activation patterns in the scapular positioning muscles during loaded scapular plane elevation: the effects of latent myofascial trigger points. Clin Biomech 25(8):765–770
179. Mokaya F, Noh HY, Lucas R, Zhang P (2018) Myovibe: enabling inertial sensor-based muscle activation detection in high-mobility exercise environments. ACM Trans Sens Netw 14(1):1638
180. Baca A, Kornfeind P (2006) Rapid feedback systems for elite sports training. IEEE Pervasive Comput 5(4):70–76
181. Schaffert N, Mattes K, Effenberg A (2011) An investigation of online acoustic information for elite rowers in on-water training conditions. J Hum Sport Exerc 6(2):392–405. https://doi.org/10.4100/jhse.2011.62.20
182. Dowling AV, Favre J, Andriacchi TP (2012) Inertial sensor-based feedback can reduce key risk metrics for anterior cruciate ligament injury during jump landings. Am J Sports Med 40(5):1075–1083. https://doi.org/10.1177/0363546512437529
183. Schwenk M, Grewal GS, Honarvar B, Schwenk S, Mohler J, Khalsa DS, Najafi B (2014) Interactive balance training integrating sensor-based visual feedback of movement performance: a pilot study in older adults. J Neuroeng Rehabil 11(1):1–13
184. Patel S, Park H, Bonato P, Chan L, Rodgers M (2012) A review of wearable sensors and systems with application in rehabilitation. J NeuroEng Rehabil 9(1):21. https://doi.org/10.1186/1743-0003-9-21
185. Harari GM, Lane N, Wang R, Crosier BS, Campbell A, Gosling S (2016) Using smartphones to collect behavioral data in psychological science. Perspect Psychol Sci 11:838–854
186. Baghdadi A (2019) Application of inertial measurement unit (IMU) in advanced human health and safety surveillance: a data fusion and machine learning approach. PhD thesis, University at Buffalo, The State University of New York
187. https://www.hopkinsmedicine.org/health/conditions-and-diseases/vital-signs-body-temperature-pulse-rate-respiration-rate-blood-pressure
188. https://corporate.epson/en/technology/search-by-products/microdevice/vital-sensing.html
189. Sun W, Guo Z, Yang Z, Wu Y, Lan W, Liao Y, Wu X, Liu Y (2022) A review of recent advances in vital signals monitoring of sports and health via flexible wearable sensors. Sensors 22(20):7784
190. De Fazio R, Stabile M, De Vittorio M, Velázquez R, Visconti P (2021) An overview of wearable piezoresistive and inertial sensors for respiration rate monitoring. Electronics 10(17):2178
191. Milici S, Lázaro A, Villarino R, Girbau D, Magnarosa M (2018) Wireless wearable magnetometer-based sensor for sleep quality monitoring. IEEE Sens J 18(5):2145–2152
192. Kebe M, Gadhafi R, Mohammad B, Sanduleanu M, Saleh H, Al-Qutayri M (2020) Human vital signs detection methods and potential using radars: a review. Sensors 20(5):1454
193. Cismas A, Cismas E, Popescu D (2019) Detecting vital signs using inertial measurement unit sensors in military applications. In: 2019 22nd international conference on control systems and computer science (CSCS). IEEE, pp 86–90
194. Kuschan J, Schmidt H, Krüger J (2017) Analysis of ergonomic and unergonomic human lifting behaviors by using inertial measurement units. Curr Dir Biomed Eng 3(1):7–10
195. Mudiyanselage SE, Nguyen PHD, Rajabi MS, Akhavian R (2021) Automated workers' ergonomic risk assessment in manual material handling using sEMG wearable sensors and machine learning. Electronics 10(20):2558
196. Humadi A, Nazarahari M, Ahmad R, Rouhani H (2021) In-field instrumented ergonomic risk assessment: inertial measurement units versus kinect V2. Int J Ind Ergon 84:103147
197. Fauziah A, Muslim K, Chandra SD (2021) Development of a real-time ergonomic assessment tool to minimize musculoskeletal disorders risk. In: Proceedings of the 2nd Asia Pacific international conference on industrial engineering and operations management, Surakarta, Indonesia. IEOM Society International
198. Hsu C-F, Lin T-T (2019) Development of an ergonomic evaluation system based on inertial measurement unit and its application for exoskeleton load reduction. In: 2019 ASABE annual international meeting. American Society of Agricultural and Biological Engineers, p 1
199. Humadi A, Nazarahari M, Ahmad R, Rouhani H (2020) Instrumented ergonomic risk assessment using wearable inertial measurement units: impact of joint angle convention. IEEE Access 9:7293–7305
200. Jahanian O, Van Straaten MG, Goodwin BM, Cain SM, Lennon RJ, Barlow JD, Murthy NS, Morrow MM (2021) Inertial measurement unit-derived ergonomic metrics for assessing arm use in manual wheelchair users with spinal cord injury: A preliminary report. Top Spinal Cord Inj Rehabil 27(3):12–25
201. Vignais N, Bernard F, Touvenot G, Sagot J-C (2017) Physical risk factors identification based on body sensor network combined to videotaping. Appl Ergon 65:410–417
202. Villalobos A, Mac Cawley A (2022) Prediction of slaughterhouse workers' RULA scores and knife edge using low-cost inertial measurement sensor units and machine learning algorithms. Appl Ergon 98:103556
203. Zhang H, Chen L, Zhang Y, Hu R, He C, Tan Y, Zhang J (2022) A wearable real-time character recognition system based on edge computing-enabled deep learning for air-writing. J Sens. https://doi.org/10.1155/2022/8507706
204. Zhao H, Ma Y, Wang S, Watson A, Zhou G (2018) Mobigesture: mobility-aware hand gesture recognition for healthcare. Smart Health 9:129–143
205. Shin S, Sung W (2016) Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks. In: 2016 IEEE international symposium on circuits and systems (ISCAS). IEEE, pp 2274–2277
206. Zhang C, Xue Q, Waghmare A, Meng R, Jain S, Han Y, Li X, Cunefare K, Ploetz T, Starner T et al (2018) 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, pp 1–10
207. Banos O, Calatroni A, Damas M, Pomares H, Roggen D, Rojas I, Villalonga C (2021) Opportunistic activity recognition in IoT sensor ecosystems via multimodal transfer learning. Neural Process Lett 59:1–29
208. Malibari AA, Alzahrani JS, Qahmash A, Maray M, Alghamdi M, Alshahrani R, Mohamed A, Hilal AM (2022) Quantum water strider algorithm with hybrid-deep-learning-based activity recognition for human–computer interaction. Appl Sci 12(14):6848
209. Basterretxea K, Echanobe J, del Campo I (2014) A wearable human activity recognition system on a chip. In: Proceedings of the 2014 conference on design and architectures for signal and image processing. IEEE, pp 1–8
210. Czabke A, Marsch S, Lueth TC (2011) Accelerometer based real-time activity analysis on a microcontroller. In: 2011 5th International conference on pervasive computing technologies

for healthcare (PervasiveHealth) and workshops. IEEE, pp 40–46
211. Li X, Zhang Y, Marsic I, Sarcevic A, Burd RS (2016) Deep learning for RFID-based activity recognition. In: Proceedings of the 14th ACM conference on embedded network sensor systems CD-ROM. ACM, pp 164–175
212. Sundaramoorthy P, Gudur GK, Moorthy MR, Bhandari RN, Vijayaraghavan V (2018) HARNet: towards on-device incremental learning using deep ensembles on constrained devices. In: Proceedings of the 2nd international workshop on embedded and mobile deep learning. ACM, pp 31–36
213. Saponas TS, Meyers BR, Brush AB (2011) Human-activity recognition (HAR) everywhere: embedded, mobile, desktop, home and cloud. In: Pervasive 2011 workshop: frontiers in activity recognition using pervasive sensing (IWFAR), microsoft research. Citeseer
214. Chen R, Luo H, Zhao F, Meng X, Xie Z, Zhu Y (2021) Modeling accurate human activity recognition for embedded devices using multi-level distillation. CoRR arXiv:abs/2107.07331
215. Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V et al (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 1314–1324
216. Bhat G, Deb R, Chaurasia VV, Shill H, Ogras UY (2018) Online human activity recognition using low-power wearable devices. In: 2018 IEEE/ACM international conference on computer-aided design (ICCAD). IEEE, pp 1–8
217. Bhat G, Deb R, Ogras UY (2019) OpenHealth: open source platform for wearable health monitoring. IEEE Des Test 36(5):27–34
218. Alessandrini M, Biagetti G, Crippa P, Falaschetti L, Turchetti C (2021) Recurrent neural network for human activity recognition in embedded systems using ppg and accelerometer data. Electronics 10(14):1715
219. Choudhury T, Borriello G, Consolvo S, Haehnel D, Harrison B, Hemingway B, Hightower J, Pedja P, Koscher K, LaMarca A et al (2008) The mobile sensing platform: an embedded activity recognition system. IEEE Pervasive Comput 7(2):32–41
220. Ravi D, Wong C, Lo B, Yang G-Z (2016) 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). IEEE, pp 71–76
221. Yagis E, Citi L, Diciotti S, Marzi C, Atnafu SW, De Herrera AGS (2020) 3D convolutional neural networks for diagnosis of Alzheimer's disease via structural MRI. In: 2020 IEEE 33rd international symposium on computer-based medical systems (CBMS). IEEE, pp 65–70
222. Wang S, Wang H, Shen Y, Wang X (2018) Automatic recognition of mild cognitive impairment and Alzheimers disease using ensemble based 3D densely connected convolutional networks. In: 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 517–523
223. Wang H, Shen Y, Wang S, Xiao T, Deng L, Wang X, Zhao X (2019) Ensemble of 3D densely connected convolutional network for diagnosis of mild cognitive impairment and Alzheimer's disease. Neurocomputing 333:145–156
224. Yu W, Lei B, Ng MK, Cheung AC, Shen Y, Wang S (2021) Tensorizing GAN with high-order pooling for Alzheimer's disease assessment. IEEE Trans Neural Netw Learn Syst 33(9):4945–4959
225. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R (2013) Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199
226. Strauss T, Hanselmann M, Junginger A, Ulmer H (2017) Ensemble methods as a defense to adversarial perturbations against deep neural networks. arXiv preprint arXiv:1709.03423
227. Karim F, Majumdar S, Darabi H (2021) Adversarial attacks on time series. IEEE Trans Pattern Anal Mach Intell 43(10):3309–3320. https://doi.org/10.1109/TPAMI.2020.2986319
228. Fawaz HI, Forestier G, Weber J, Idoumghar L, Muller P (2019) Adversarial attacks on deep neural networks for time series classification. In: 2019 International joint conference on neural networks (IJCNN), pp 1–8. https://doi.org/10.1109/IJCNN.2019.8851936
229. https://hangyuan.xyz/2022/11/10/why_human_activity_recognition_is_far_from_solved.html
230. Attig C, Franke T (2020) Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Hum Behav 102:223–237
231. Pham C, Olivier P (2009) Slice & dice: Recognizing food preparation activities using embedded accelerometers. In: European conference on ambient intelligence. Springer, pp 34–43
232. De la Torre F, Hodgins J, Bargteil A, Martin X, Macey J, Collado A, Beltran P (2008) Guide to the Carnegie Mellon university multimodal activity (CMU-MMAC) database. Robotics Institute, 135
233. Huang C, Zhang F, Xu Z, Wei J (2022) The diverse gait dataset: gait segmentation using inertial sensors for pedestrian localization with different genders, heights and walking speeds. Sensors 22(4):1678
234. Mäkela S-M, Lämsä A, Keränen JS, Liikka J, Ronkainen J, Peltola J, Häikiö J, Järvinen S, Bordallo López M (2021) Introducing VTT-coniot: a realistic dataset for activity recognition of construction workers using IMU devices. Sustainability 14(1):220
235. Alemayoh TT, Lee JH, Okamoto S (2021) New sensor data structuring for deeper feature extraction in human activity recognition. Sensors 21(8):2814
236. Zhou Q, Shan J, Fang B, Zhang S, Sun F, Ding W, Wang C, Zhang Q (2021) Personal-specific gait recognition based on latent orthogonal feature space. Cogn Comput Syst 3(1):61–69
237. Kim H, Kim H-J, Park J, Ryu J-K, Kim S-C (2021) Recognition of fine-grained walking patterns using a smartwatch with deep attentive neural networks. Sensors 21(19):6393
238. Li R, Balakrishnan GP, Nie J, Li Y, Agu E, Grimone K, Herman D, Abrantes AM, Stein MD (2021) Estimation of blood alcohol concentration from smartphone gait data using neural networks. IEEE Access 9:61237–61255
239. Madcor H, Alnaggar A, Rohaim Y, Osman M, Mostafa M, Adel O, Gomaa W (2021) Vsgd: a bi-modal dataset for gait analysis. In: 2021 International mobile, intelligent, and ubiquitous computing conference (MIUCC). IEEE, pp 313–318
240. Fuller D, Anaraki JR, Simango B, Rayner M, Dorani F, Bozorgi A, Luan H, Basset FA (2021) Predicting lying, sitting, walking and running using apple watch and fitbit data. BMJ Open Sport Exerc Med 7(1):001004
241. Fang B, Zhou Q, Sun F, Shan J, Wang M, Xiang C, Zhang Q (2020) Gait neural network for human–exoskeleton interaction. Front Neurorobot 14:58
242. Zhang W, Zhao X, Li Z (2019) A comprehensive study of smartphone-based indoor activity recognition via xgboost. IEEE Access 7:80027–80042
243. He J, Zhang Q, Wang L, Pei L (2018) Weakly supervised human activity recognition from wearable sensors by recurrent attention learning. IEEE Sens J 19(6):2287–2297

244. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H (2018) Protecting sensory data against sensitive inferences. In: Proceedings of the 1st workshop on privacy by design in distributed systems W-P2DS'18. ACM, pp 1–6
245. Genovese V, Mannini A, Sabatini AM (2017) A smartwatch step counter for slow and intermittent ambulation. IEEE Access 5:13028–13037
246. Khandelwal S, Wickström N (2017) Evaluation of the performance of accelerometer-based gait event detection algorithms in different real-world scenarios using the MAREA gait database. Gait Posture 51:84–90
247. Esfahani P, Malazi HT (2017) PAMS: a new position-aware multi-sensor dataset for human activity recognition using smartphones. In: 2017 19th International symposium on computer architecture and digital systems (CADS). IEEE, pp 1–7
248. Dutta A, Ma O, Toledo M, Buman MP, Bliss DW (2016) Comparing Gaussian mixture model and hidden Markov model to classify unique physical activities from accelerometer sensor data. In: 2016 15th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 339–346
249. Bruno B, Mastrogiovanni F, Sgorbissa A (2015) Wearable inertial sensors: applications, challenges, and public test benches. IEEE Robot Autom Mag 22(3):116–124
250. Gupta S, Kumar A (2015) Human activity recognition through smartphone's tri-axial accelerometer using time domain wave analysis and machine learning. Int J Comput Appl 975:8887
251. Bayat A, Pomplun M, Tran DA (2014) A study on human activity recognition using accelerometer data from smartphones. Procedia Comput Sci 34:450–457
252. Aung MS, Thies SB, Kenney LP, Howard D, Selles RW, Findlow AH, Goulermas JY (2013) Automated detection of instantaneous gait events using time frequency analysis and manifold embedding. IEEE Trans Neural Syst Rehabil Eng 21(6):908–916
253. Casale P, Pujol O, Radeva P (2012) Personalization and user verification in wearable systems using biometric walking patterns. Pers Ubiquitous Comput 16(5):563–580
254. Lara OD, Labrador MA (2012) A mobile platform for real-time human activity recognition. In: 2012 IEEE consumer communications and networking conference (CCNC). IEEE, pp 667–671
255. Lara OD, Pérez AJ, Labrador MA, Posada JD (2012) Centinela: a human activity recognition system based on acceleration and vital sign data. Pervasive Mob Comput 8(5):717–729
256. Bovi G, Rabuffetti M, Mazzoleni P, Ferrarin M (2011) A multiple-task gait analysis approach: kinematic, kinetic and EMG reference data for healthy young and adult subjects. Gait Posture 33(1):6–13
257. Longstaff B, Reddy S, Estrin D (2010) Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In: 2010 4th International conference on pervasive computing technologies for healthcare. IEEE, pp 1–7
258. Santos AC, Cardoso JM, Ferreira DR, Diniz PC, Chaínho P (2010) Providing user context for mobile and social networking applications. Pervasive Mob Comput 6(3):324–341
259. Krishnan NC, Colbry D, Juillard C, Panchanathan S (2008) Real time human activity recognition using tri-axial accelerometers. In: Sensors, signals and information processing workshop, vol 2008. Multidisciplinary Digital Publishing Institute, p. 3337–3340
260. Wang N, Ambikairajah E, Lovell NH, Celler BG (2007) Accelerometry based classification of walking patterns using time-frequency analysis. In: 2007 29th Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 4899–4902
261. Coley B, Najafi B, Paraschiv-Ionescu A, Aminian K (2005) Stair climbing detection during daily physical activity using a miniature gyroscope. Gait Posture 22(4):287–294
262. Lee S-W, Mase K (2002) Activity and location recognition using wearable sensors. IEEE Pervasive Comput 1(3):24–32
263. Randell C, Muller H (2000) Context awareness by analysing accelerometer data. In: Digest of papers. Fourth international symposium on wearable computers. IEEE, pp 175–176
264. Biagetti G, Crippa P, Falaschetti L, Saraceni L, Tiranti A, Turchetti C (2020) Dataset from PPG wireless sensor for activity monitoring. Data Brief 29:105044
265. Lee J, Joo H, Lee J, Chee Y (2020) Automatic classification of squat posture using inertial sensors: deep learning approach. Sensors 20(2):361
266. Tuncer T, Ertam F, Dogan S, Aydemir E, Pławiak P (2020) Ensemble residual network-based gender and activity recognition method with signals. J Supercomput 76(3):2119–2138
267. Zhang M, Chen S, Zhao X, Yang Z (2018) Research on construction workers' activity recognition based on smartphone. Sensors 18(8):2667
268. Calvo AF, Holguin GA, Medeiros H (2018) Human activity recognition using multi-modal data fusion. In: Iberoamerican congress on pattern recognition. Springer, pp 946–953
269. Koskimäki H, Siirtola P, Röning J (2017) Myogym: introducing an open gym data set for activity recognition collected using myo armband. In: Proceedings of the 2017 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 ACM international symposium on wearable computers, pp 537–546
270. Liu G, Liang J, Lan G, Hao Q, Chen M (2016) Convolution neutral network enhanced binary sensor network for human activity recognition. In: 2016 IEEE sensors. IEEE, pp 1–3
271. Lim T-K, Park S-M, Lee H-C, Lee D-E (2016) Artificial neural network-based slip-trip classifier using smart sensor for construction workplace. J Constr Eng Manag 142(2):04015065
272. Chen C, Jafari R, Kehtarnavaz N (2015) 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). IEEE, pp 168–172
273. Mitchell E, Monaghan D, O'Connor NE (2013) Classification of sporting activities using smartphone accelerometers. Sensors 13(4):5317–5337
274. Bayati H, Mill JdR, Chavarriaga R et al (2011) Unsupervised adaptation to on-body sensor displacement in acceleration-based activity recognition. In: 2011 15th Annual international symposium on wearable computers. IEEE, pp 71–78
275. Liu S, Gao RX, John D, Staudenmayer JW, Freedson PS (2011) Multisensor data fusion for physical activity assessment. IEEE Trans Biomed Eng 59(3):687–696
276. Tunçel O, Altun K, Barshan B (2009) Classifying human leg motions with uniaxial piezoelectric gyroscopes. Sensors 9(11):8508–8546
277. Ermes M, Pärkkä J, Mäntyjärvi J, Korhonen I (2008) Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans Inf Technol Biomed 12(1):20–26
278. Tapia EM, Intille SS, Haskell W, Larson K, Wright J, King A, Friedman R (2007) Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: 11th IEEE international symposium on wearable computers, 2007. IEEE, pp 37–40
279. Olguın DO, Pentland AS (2006) Human activity recognition: accuracy across common locations for wearable sensors. In: Proceedings of 2006 10th IEEE international symposium on wearable computers, Montreux, Switzerland. Citeseer, pp 11–14

280. Chambers GS, Venkatesh S, West GA, Bui HH (2002) Hierarchical recognition of intentional human gestures for sports video annotation. In: Object recognition supported by user interaction for service robots, vol 2. IEEE, pp 1082–1085
281. Herren R, Sparti A, Aminian K, Schutz Y (1999) The prediction of speed and incline in outdoor running in humans using accelerometry. Med Sci Sports Exerc 31(7):1053–1059
282. Xia Q, Korpela J, Namioka Y, Maekawa T (2020) Robust unsupervised factory activity recognition with body-worn accelerometer using temporal structure of multiple sensor data motifs. Proc ACM Interact Mob Wearable Ubiquitous Technol 4(3):1–30
283. Akhavian R, Behzadan A (2015) Wearable sensor-based activity recognition for data-driven simulation of construction workers' activities. In: 2015 Winter simulation conference (WSC). IEEE, pp 3333–3344
284. Chifu VR, Pop CB, Demjen D, Socaci R, Todea D, Antal M, Cioara T, Anghel I, Antal C (2022) Identifying and monitoring the daily routine of seniors living at home. Sensors 22(3):992
285. Zhang L, Zhu Y, Jiang M, Wu Y, Deng K, Ni Q (2021) Body temperature monitoring for regular COVID-19 prevention based on human daily activity recognition. Sensors 21(22):7540
286. Lakoju M, Ajienka N, Khanesar MA, Burnap P, Branson DT (2021) Unsupervised learning for product use activity recognition: an exploratory study of a "chatty device''. Sensors 21(15):4991
287. de Sousa FASF, Escriba C, Bravo EGA, Brossa V, Fourniols J-Y, Rossi C (2021) Wearable pre-impact fall detection system based on 3D accelerometer and subject's height. IEEE Sens J 22(2):1738–1745
288. Skoglund MA, Balzi G, Jensen EL, Bhuiyan TA, Rotger-Griful S (2021) Activity tracking using ear-level accelerometers. Front Digit Health 122:724714
289. Gochoo M, Tahir SBUD, Jalal A, Kim K (2021) Monitoring real-time personal locomotion behaviors over smart indoor–outdoor environments via body-worn sensors. IEEE Access 9:70556–70570
290. Choudhury NA, Moulik S, Roy DS. HARSense: statistical human activity recognition dataset. IEEE Dataport. https://doi.org/10.21227/9pt3-2m34
291. Palmerini L, Klenk J, Becker C, Chiari L (2020) Accelerometer-based fall detection using machine learning: training and testing on real-world falls. Sensors 20(22):6479
292. Li Q, Yang Y, Yang P (2020) Human activity recognition based on triaxial accelerometer using multi-feature weighted ensemble. In: 2020 IEEE 18th international conference on industrial informatics (INDIN), vol 1. IEEE, pp 561–566
293. Bhat G, Tran N, Shill H, Ogras UY (2020) w-HAR: an activity recognition dataset and framework using low-power wearable devices. Sensors 20(18):5356
294. Wang L, Gjoreski H, Ciliberto M, Lago P, Murao K, Okita T, Roggen D (2020) Summary of the Sussex–Huawei locomotion-transportation recognition challenge 2020. 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, pp 351–358
295. Garcia-Gonzalez D, Rivero D, Fernandez-Blanco E, Luaces MR (2020) A public domain dataset for real-life human activity recognition using smartphone sensors. Sensors 20(8):2200
296. Ebner M, Fetzer T, Bullmann M, Deinzer F, Grzegorzek M (2020) Recognition of typical locomotion activities based on the sensor data of a smartphone in pocket or hand. Sensors 20(22):6559
297. Alves J, Silva J, Grifo E, Resende C, Sousa I (2019) Wearable embedded intelligence for detection of falls independently of on-body location. Sensors 19(11):2426
298. Weiss GM, Yoneda K, Hayajneh T (2019) Smartphone and smartwatch-based biometrics using activities of daily living. IEEE Access 7:133190–133202
299. Balli S, Sağbaş EA, Peker M (2019) Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. Meas Control 52(1–2):37–45
300. Reiss A, Indlekofer I, Schmidt P, Van Laerhoven K (2019) Deep PPG: large-scale heart rate estimation with convolutional neural networks. Sensors 19(14):3079
301. Ni Q, Zhang L, Li L (2018) A heterogeneous ensemble approach for activity recognition with integration of change point-based data segmentation. Appl Sci 8(9):1695
302. Liu K-C, Hsieh C-Y, Chan C-T (2018) Transition-aware housekeeping task monitoring using single wrist-worn sensor. IEEE Sens J 18(21):8950–8962
303. Micucci D, Mobilio M, Napoletano P (2017) Unimib SHAR: a dataset for human activity recognition using acceleration data from smartphones. Appl Sci. https://doi.org/10.3390/app7101101
304. Weiss GM, Timko JL, Gallagher CM, Yoneda K, Schreiber AJ (2016) Smartwatch-based activity recognition: a machine learning approach. In: 2016 IEEE-EMBS international conference on biomedical and health informatics (BHI). IEEE, pp 426–429
305. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ (2016) Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors 16(4):426
306. Twomey N, Diethe T, Kull M, Song H, Camplani M, Hannuna S, Fafoutis X, Zhu N, Woznowski P, Flach P et al (2016) The SPHERE challenge: activity recognition with multimodal sensor data. arXiv preprint arXiv:1603.00797
307. Davis K, Owusu E, Bastani V, Marcenaro L, Hu J, Regazzoni C, Feijs L (2016) Activity recognition based on inertial sensors for ambient assisted living. In: 2016 19th International conference on information fusion (fusion). IEEE, pp 371–378
308. Piyare RK, Lee SR (2014) Activity recognition of workers and passengers onboard ships using multimodal sensors in a smartphone. J Korean Inst Commun Inf Sci 39(9):811–819
309. Lockhart JW, Weiss GM (2014) The benefits of personalized smartphone-based activity recognition models. In: Proceedings of the 2014 SIAM international conference on data mining. SIAM, pp 614–622
310. Khan AM, Tufail A, Khattak AM, Laine TH (2014) Activity recognition on smartphones via sensor-fusion and KDA-based SVMs. Int J Distrib Sens Netw 10(5):503291
311. Siirtola P, Röning J (2013) Ready-to-use activity recognition for smartphones. In: 2013 IEEE symposium on computational intelligence and data mining (CIDM). IEEE, pp 59–64
312. Cleland I, Kikhia B, Nugent C, Boytsov A, Hallberg J, Synnes K, McClean S, Finlay D (2013) Optimal placement of accelerometers for the detection of everyday activities. Sensors 13(7):9183–9200
313. Prudêncio J, Aguiar A, Lucani D (2013) Physical activity recognition from smartphone embedded sensors. In: Iberian conference on pattern recognition and image analysis. Springer, pp 863–872
314. Trabelsi D, Mohammed S, Chamroukhi F, Oukhellou L, Amirat Y (2013) An unsupervised approach for automatic activity recognition based on hidden Markov model regression. IEEE Trans Autom Sci Eng 10(3):829–835
315. Thiemjarus S, Henpraserttae A, Marukatat S (2013) A study on instance-based learning with reduced training prototypes for

device-context-independent activity recognition on a mobile phone. In: 2013 IEEE international conference on body sensor networks. IEEE, pp 1–6
316. Leutheuser H, Schuldhaus D, Eskofier BM (2013) Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. PLoS ONE 8(10):75196
317. Ofli F, Chaudhry R, Kurillo G, Vidal R, Bajcsy R (2013) Berkeley MHAD: a comprehensive multimodal human action database. In: 2013 IEEE workshop on applications of computer vision (WACV). IEEE, pp 53–60
318. Liang Y, Zhou X, Yu Z, Guo B, Yang Y (2012) Energy efficient activity recognition based on low resolution accelerometer in smart phones. In: International conference on grid and pervasive computing. Springer, pp 122–136
319. Ugulino W, Cardador D, Vega K, Velloso E, Milidiú R, Fuks H (2012) Wearable computing: accelerometers' data classification of body postures and movements. In: Brazilian symposium on artificial intelligence. Springer, pp 52–61
320. Albert MV, Toledo S, Shapiro M, Koerding K (2012) Using mobile phones for activity recognition in Parkinson's patients. Front Neurol 3:158
321. Hattori Y, Inoue S (2012) A large scale gathering system for activity data using mobile devices. Inf Media Technol 7(1):458–465
322. Dernbach S, Das B, Krishnan NC, Thomas BL, Cook DJ (2012) Simple and complex activity recognition through smart phones. In: 2012 IEEE eighth international conference on intelligent environments. IEEE, pp 214–221
323. Schindhelm CK (2012) Activity recognition and step detection with smartphones: towards terminal based indoor positioning system. In: 2012 IEEE 23rd international symposium on personal, indoor and mobile radio communications-(PIMRC). IEEE, pp 2454–2459
324. Chuang F-C, Wang J-S, Yang Y-T, Kao T-P (2012) A wearable activity sensor system and its physical activity classification scheme. In: The 2012 international joint conference on neural networks (IJCNN). IEEE, pp 1–6
325. Ioana-Iuliana F, Rodica-Elena D (2011) Detection of daily movements from data collected with two tri-axial accelerometers. In: 2011 34th International conference on telecommunications and signal processing (TSP). IEEE, pp 376–380
326. Chernbumroong S, Atkins AS, Yu H (2011) Activity classification using a single wrist-worn accelerometer. In: 2011 5th International conference on software, knowledge information, industrial management and applications (SKIMA) proceedings. IEEE, pp 1–6
327. Wang F, Wang M, Feng N (2011) Research on classification of human daily activities based on a single tri-axial accelerometer. In: 2011 First international workshop on complexity and data mining. IEEE, pp 121–124
328. Gjoreski H, Lustrek M, Gams M (2011) Accelerometer placement for posture recognition and fall detection. In: 2011 Seventh international conference on intelligent environments. IEEE, pp 47–54
329. Zhang Y, Markovic S, Sapir I, Wagenaar RC, Little TD (2011) Continuous functional activity monitoring based on wearable tri-axial accelerometer and gyroscope. In: 2011 5th International conference on pervasive computing technologies for healthcare (PervasiveHealth) and workshops. IEEE, pp 370–373
330. Min J-K, Cho S-B (2011) Activity recognition based on wearable sensors using selection/fusion hybrid ensemble. In: 2011 IEEE international conference on systems, man, and cybernetics. IEEE, pp 1319–1324
331. Xu M, Goldfain A, Chowdhury AR, DelloStritto J (2011) Towards accelerometry based static posture identification. In: 2011 IEEE consumer communications and networking conference (CCNC). IEEE, pp 29–33
332. Atallah L, Lo B, King R, Yang G-Z (2011) Sensor positioning for activity recognition using wearable accelerometers. IEEE Trans Biomed Circuits Syst 5(4):320–329
333. Andreu J, Baruah RD, Angelov P (2011) Real time recognition of human activities from wearable sensors by evolving classifiers. In: 2011 IEEE international conference on fuzzy systems (FUZZ-IEEE 2011). IEEE, pp 2786–2793
334. Martín H, Bernardos AM, Tarrío P, Casar JR (2011) Enhancing activity recognition by fusing inertial and biometric information. In: 14th International conference on information fusion. IEEE, pp 1–8
335. Riboni D, Bettini C (2011) Cosar: hybrid reasoning for context-aware activity recognition. Pers Ubiquitous Comput 15(3):271–289
336. Maekawa T, Watanabe S (2011) Unsupervised activity recognition with user's physical characteristics data. In: 2011 15th annual international symposium on wearable computers. IEEE, pp 89–96
337. Lee M-W, Khan AM, Kim T-S (2011) A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. Pers Ubiquitous Comput 15(8):887–898
338. Alvarez-Alvarez A, Trivino G, Cordón O (2011) Body posture recognition by means of a genetic fuzzy finite state machine. In: 2011 IEEE 5th international workshop on genetic and evolutionary fuzzy systems (GEFS). IEEE, pp 60–65
339. Reiss A, Stricker D (2011) Introducing a modular activity monitoring system. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 5621–5624
340. Kawaguchi N, Ogawa N, Iwasaki Y, Kaji K, Terada T, Murao K, Inoue S, Kawahara Y, Sumi Y, Nishio N (2011) HASC challenge: gathering large scale human activity corpus for the real-world activity understandings. In: Proceedings of the 2nd augmented human international conference. ACM, p 27
341. Gao L, Bourke AK, Nelson J (2011) A system for activity recognition using multi-sensor fusion. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 7869–7872
342. Lee Y-S, Cho S-B (2011) Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer. In: International conference on hybrid artificial intelligence systems. Springer, pp 460–467
343. Hong J-H, Yang S-I, Cho S-B (2010) Conamsn: a context-aware messenger using dynamic Bayesian networks with wearable sensors. Expert Syst Appl 37(6):4680–4686
344. Lu H, Yang J, Liu Z, Lane ND, Choudhury T, Campbell AT (2010) The jigsaw continuous sensing engine for mobile phone applications. In: Proceedings of the 8th ACM conference on embedded networked sensor systems. ACM, pp 71–84
345. Frank K, Vera Nadales MJ, Robertson P, Pfeifer T (2010) Bayesian recognition of motion related activities with inertial sensors. In: Proceedings of the 12th ACM international conference adjunct papers on ubiquitous computing-adjunct. ACM, pp 445–446
346. Van Kasteren T, Englebienne G, Kröse BJ (2010) Activity recognition using semi-Markov models on real world smart home datasets. J Ambient Intell Smart Environ 2(3):311–325
347. Khan AM, Lee YK, Lee SY (2010) Accelerometer's position free human activity recognition using a hierarchical recognition model. In: The 12th IEEE international conference on e-health networking, applications and services. IEEE, pp 296–301

348. Sazonov ES, Fulk G, Hill J, Schutz Y, Browning R (2010) Monitoring of posture allocations and activities by a shoe-based wearable sensor. IEEE Trans Biomed Eng 58(4):983–990
349. Gu T, Wu Z, Wang L, Tao X, Lu J (2009) Mining emerging patterns for recognizing activities of multiple users in pervasive computing. In: 2009 6th Annual international mobile and ubiquitous systems: networking and services, MobiQuitous. IEEE, pp 1–10
350. Roy SH, Cheng MS, Chang S-S, Moore J, De Luca G, Nawab SH, De Luca CJ (2009) A combined sEMG and accelerometer system for monitoring functional activity in stroke. IEEE Trans Neural Syst Rehabil Eng 17(6):585–594
351. Bonomi AG, Goris A, Yin B, Westerterp KR (2009) Detection of type, duration, and intensity of physical activity using an accelerometer. Med Sci Sports Exerc (MSSE) 41(9):1770–1777
352. Song S-k, Jang J, Park S-J (2009) Dynamic activity classification based on automatic adaptation of postural orientation. In: 2009 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 6175–6178
353. Maguire D, Frisby R (2009) Comparison of feature classification algorithms for activity recognition based on accelerometer and heart rate data. In: 9th IT &T conference, p 11
354. Győrbíró N, Fábián Á, Hományi G (2009) An activity recognition system for mobile phones. Mob Netw Appl 14(1):82–91
355. Guenterberg E, Ghasemzadeh H, Loseu V, Jafari R (2009) Distributed continuous action recognition using a hidden Markov model in body sensor networks. In: International conference on distributed computing in sensor systems. Springer, pp 145–158
356. Yang J-Y, Wang J-S, Chen Y-P (2008) Using acceleration measurements for activity recognition: an effective learning algorithm for constructing neural classifiers. Pattern Recognit Lett 29(16):2213–2220
357. Krishnan NC, Panchanathan S (2008) Analysis of low resolution accelerometer data for continuous human activity recognition. In: 2008 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 3337–3340
358. Hong Y-J, Kim I-J, Ahn SC, Kim H-G (2008) Activity recognition using wearable sensors for elder care. In: 2008 Second international conference on future generation communication and networking, vol 2. IEEE, pp 302–305
359. Song S, Jang J, Park S (2008) A phone for human activity recognition using triaxial acceleration sensor. In: 2008 Digest of technical papers-international conference on consumer electronics. IEEE, pp 1–2
360. Yeoh W-S, Pek I, Yong Y-H, Chen X, Waluyo AB (2008) Ambulatory monitoring of human posture and walking speed using wearable accelerometer sensors. In: 2008 30th annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 5184–5187
361. Van Kasteren T, Noulas A, Englebienne G, Kröse B (2008) Accurate activity recognition in a home setting. In: Proceedings of the 10th international conference on ubiquitous computing. ACM, pp 1–9
362. Huynh T, Fritz M, Schiele B (2008) Discovery of activity patterns using topic models. In: Proceedings of the 10th international conference on ubiquitous computing, pp 10–19
363. Lo B, Atallah L, Aziz O, ElHew ME, Darzi A, Yang G-Z (2007) Real-time pervasive monitoring for postoperative care. In: 4th International workshop on wearable and implantable body sensor networks (BSN 2007). IEEE, pp 122–127
364. Maurer U, Smailagic A, Siewiorek DP, Deisher M (2006) Activity recognition and monitoring using multiple sensors on different body positions. In: International workshop on wearable and implantable body sensor networks (BSN 2006). IEEE, p 4
365. Pirttikangas S, Fujinami K, Nakajima T (2006) Feature selection and activity recognition from wearable sensors. In: International symposium on ubiquitious computing systems. Springer, pp 516–527
366. Allen FR, Ambikairajah E, Lovell NH, Celler BG (2006) Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiol Meas 27(10):935
367. Blum M, Pentland A, Troster G (2006) Insense: interest-based life logging. IEEE MultiMedia 13(4):40–48
368. Tapia EM, Intille SS, Lopez L, Larson K (2006) The design of a portable kit of wireless sensors for naturalistic data collection. In: International conference on pervasive computing. Springer, pp 117–134
369. Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I (2006) Activity classification using realistic data from wearable sensors. IEEE Trans Inf Technol Biomed 10(1):119–128
370. Ganti RK, Jayachandran P, Abdelzaher TF, Stankovic JA (2006) Satire: a software architecture for smart attire. In: Proceedings of the 4th international conference on mobile systems, applications and services. ACM, pp 110–123
371. Karantonis DM, Narayanan MR, Mathie M, Lovell NH, Celler BG (2006) Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. IEEE Trans Inf Technol Biomed 10(1):156–167
372. Lyons G, Culhane K, Hilton D, Grace P, Lyons D (2005) A description of an accelerometer-based mobility monitoring technique. Med Eng Phys 27(6):497–504
373. Krause A, Smailagic A, Siewiorek DP (2005) Context-aware mobile computing: learning context-dependent personal preferences from a wearable sensor array. IEEE Trans Mob Comput 5(2):113–127
374. Ravi N, Dandekar N, Mysore P, Littman ML (2005) Activity recognition from accelerometer data. In: Proceedings of the twentieth national conference on artificial intelligence, vol 5. American Association for Artificial Intelligence, pp 1541–1546
375. Mathie MJ, Celler BG, Lovell NH, Coster AC (2004) Classification of basic daily movements using a triaxial accelerometer. Med Biol Eng Comput 42(5):679–687
376. Kern N, Schiele B, Schmidt A (2003) Multi-sensor activity context detection for wearable computing. In: European symposium on ambient intelligence. Springer, pp 220–232
377. Lee S, Park H, Hong S, Lee K, Kim Y (2003) A study on the activity classification using a triaxial accelerometer. In: Proceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439), vol 3. IEEE, pp 2941–2943
378. Van Laerhoven K, Cakmakci O (2000) What shall we teach our pants? In: Fourth international symposium on wearable computers (ISWC 2000). IEEE
379. Aminian K, Robert P, Buchser E, Rutschmann B, Hayoz D, Depairon M (1999) Physical activity monitoring based on accelerometry: validation and comparison with video observation. Med Biol Eng Comput 37(3):304–308
380. Foerster F, Smeja M, Fahrenberg J (1999) Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring. Comput Hum Behav 15(5):571–583
381. Uiterwaal M, Glerum E, Busser H, Van Lummel R (1998) Ambulatory monitoring of physical activity in working situations, a validation study. J Med Eng Technol 22(4):168–172
382. Bouten CV, Koekkoek KT, Verduin M, Kodde R, Janssen JD (1997) A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. IEEE Trans Biomed Eng 44(3):136–147

383. Mesanza AB, D'Ascanio I, Zubizarreta A, Palmerini L, Chiari L, Cabanes I (2021) Machine learning based fall detector with a sensorized tip. IEEE Access 9:164106–164117
384. Luna-Perejón F, Muñoz-Saavedra L, Civit-Masot J, Civit A, Domínguez-Morales M (2021) Ankfall-falls, falling risks and daily-life activities dataset with an ankle-placed accelerometer and training using recurrent neural networks. Sensors 21(5):1889
385. Chai X, Wu R, Pike M, Jin H, Chung W-Y, Lee B-G (2021) Smart wearables with sensor fusion for fall detection in firefighting. Sensors 21(20):6770
386. Patel W, Patel C, Ramani B, Bhaskar S, Patel M (2020) Vitafall: Nxtgeuh system for well-being monitoring with fall recognition and real-time vital sign monitoring. Int J Recent Technol Eng (IJRTE) 8(5):5139–5145
387. Martínez-Villaseñor L, Ponce H, Brieva J, Moya-Albor E, Núñez-Martínez J, Peñafort-Asturiano C (2019) Up-fall detection dataset: a multimodal approach. Sensors 19(9):1988
388. Shahzad A, Kim K (2018) Falldroid: an automated smart-phone-based fall detection system using multiple kernel learning. IEEE Trans Ind Inform 15(1):35–44
389. de Quadros T, Lazzaretti AE, Schneider FK (2018) A movement decomposition and machine learning-based fall detection system using wrist wearable device. IEEE Sens J 18(12):5082–5089
390. Saha SS, Rahman S, Rasna MJ, Islam AM, Ahad MAR (2018) DU-MD: an open-source human action dataset for ubiquitous wearable sensors. In: 2018 Joint 7th international conference on informatics, electronics and vision (ICIEV) and 2018 2nd international conference on imaging, vision and pattern recognition (icIVPR). IEEE, pp 567–572
391. Chen K-H, Yang J-J, Jaw F-S (2016) Accelerometer-based fall detection using feature extraction and support vector machine algorithms. Instrum Sci Technol 44(4):333–342
392. Casilari E, Santoyo-Ramón JA, Cano-García JM (2017) UMA-Fall: a multisensor dataset for the research on automatic fall detection. Procedia Comput Sci 110:32–39
393. Sucerquia A, López JD, Vargas-Bonilla JF (2017) Sisfall: a fall and movement dataset. Sensors 17(1):198
394. Klenk J, Schwickert L, Palmerini L, Mellone S, Bourke A, Ihlen EA, Kerse N, Hauer K, Pijnappels M, Synofzik M et al (2016) The farseeing real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls. Eur Rev Aging Phys Act 13(1):1–7
395. Wang C, Lu W, Narayanan MR, Chang DCW, Lord SR, Redmond SJ, Lovell NH (2016) Low-power fall detector using triaxial accelerometry and barometric pressure sensing. IEEE Trans Ind Inform 12(6):2302–2311
396. Vilarinho T, Farshchian B, Bajer DG, Dahl OH, Egge I, Hegdal SS, Lønes A, Slettevold JN, Weggersen SM (2015) A combined smartphone and smartwatch fall detection system. In: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing. IEEE, pp 1443–1448
397. Ojetola O, Gaura E, Brusey J (2015) Data set for fall events and daily activities from inertial sensors. In: Proceedings of the 6th ACM multimedia systems conference. ACM, pp 243–248
398. Wertner A, Czech P, Pammer-Schindler V (2015) An open labelled dataset for mobile phone sensing based fall detection. In: Proceedings of the 12th EAI international conference on mobile and ubiquitous systems: computing, networking and services. European Alliance for Innovation (EAI), pp 277–278
399. Shibuya N, Nukala BT, Rodriguez AI, Tsay J, Nguyen TQ, Zupancic S, Lie DY (2015) A real-time fall detection system using a wearable gait analysis sensor and a support vector machine (SVM) classifier. In: 2015 Eighth international conference on mobile computing and ubiquitous networking (ICMU). IEEE, pp 66–67
400. Fortino G, Gravina R (2015) Fall-mobileguard: a smart real-time fall detection system. In: Proceedings of the 10th EAI international conference on body area networks. European Alliance for Innovation (EAI), pp 44–50
401. Pierleoni P, Belli A, Palma L, Pellegrini M, Pernini L, Valenti S (2015) A high reliability wearable device for elderly fall detection. IEEE Sens J 15(8):4544–4553
402. Sabatini AM, Ligorio G, Mannini A, Genovese V, Pinna L (2015) Prior-to-and post-impact fall detection using inertial and barometric altimeter measurements. IEEE Trans Neural Syst Rehabil Eng 24(7):774–783
403. Gjoreski H, Kozina S, Gams M, Lustrek M, Álvarez-García JA, Hong J-H, Ramos J, Dey AK, Bocca M, Patwari N (2015) Competitive live evaluations of activity-recognition systems. IEEE Pervasive Comput 14(1):70–77
404. Kwolek B, Kepski M (2014) Human fall detection on embedded platform using depth maps and wireless accelerometer. Comput Methods Programs Biomed 117(3):489–501
405. Aziz O, Russell CM, Park EJ, Robinovitch SN (2014) The effect of window size and lead time on pre-impact fall detection accuracy using support vector machine analysis of waist mounted inertial sensor data. In: 2014 36th Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 30–33
406. Medrano C, Igual R, Plaza I, Castro M (2014) Detecting falls as novelties in acceleration patterns acquired with smartphones. PLoS ONE 9(4):94811
407. Vavoulas G, Pediaditis M, Chatzaki C, Spanakis EG, Tsiknakis M (2014) The Mobifall dataset: fall detection and classification with a smartphone. Int J Monit Surveil Technol Res (IJMSTR) 2(1):44–56
408. Kepski M, Kwolek B (2014) Fall detection using ceiling-mounted 3D depth camera. In: 2014 International conference on computer vision theory and applications (VISAPP), vol 2. IEEE, pp 640–647
409. Tong L, Song Q, Ge Y, Liu M (2013) Hmm-based human fall detection and prediction method using tri-axial accelerometer. IEEE Sens J 13(5):1849–1856
410. Zhao G, Mei Z, Liang D, Ivanov K, Guo Y, Wang Y, Wang L (2012) Exploration and implementation of a pre-impact fall recognition method based on an inertial body sensor network. Sensors 12(11):15338–15355
411. Kerdegari H, Samsudin K, Ramli AR, Mokaram S (2012) Evaluation of fall detection classification approaches. In: 2012 4th International conference on intelligent and advanced systems (ICIAS2012), vol 1. IEEE, pp 131–136
412. Kangas M, Vikman I, Nyberg L, Korpelainen R, Lindblom J, Jämsä T (2012) Comparison of real-life accidental falls in older people with experimental falls in middle-aged test subjects. Gait Posture 35(3):500–505
413. Lee RY, Carlisle AJ (2011) Detection of falls using accelerometers and mobile phone technology. Age Ageing 40(6):690–696
414. Ojetola O, Gaura EI, Brusey J (2011) Fall detection with wearable sensors–safe (smart fall detection). In: 2011 Seventh international conference on intelligent environments. IEEE, pp 318–321
415. Shan S, Yuan T (2010) A wearable pre-impact fall detector using feature selection and support vector machine. In: IEEE 10th international conference on signal processing proceedings. IEEE, pp 1686–1689
416. Curone D, Bertolotti GM, Cristiani A, Secco EL, Magenes G (2010) A real-time and self-calibrating algorithm based on

triaxial accelerometer signals for the detection of human posture and activity. IEEE Trans Inf Technol Biomed 14(4):1098–1105
417. Dai J, Bai X, Yang Z, Shen Z, Xuan D (2010) Perfalld: a pervasive fall detection system using mobile phones. In: 2010 8th IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops). IEEE, pp 292–297
418. Bianchi F, Redmond SJ, Narayanan MR, Cerutti S, Lovell NH (2010) Barometric pressure and triaxial accelerometry-based falls event detection. IEEE Trans Neural Syst Rehabil Eng 18(6):619–627
419. Yavuz G, Kocak M, Ergun G, Alemdar HO, Yalcin H, Incel OD, Ersoy C (2010) A smartphone based fall detector with online location support. In: Proceedings of the ACM international workshop on sensing for app phones. Association for Computing Machinery (ACM) Zurich, Switzerland, pp 31–35
420. Li Q, Stankovic JA, Hanson MA, Barth AT, Lach J, Zhou G (2009) Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. In: 2009 Sixth international workshop on wearable and implantable body sensor networks. IEEE, pp 138–143
421. Luštrek M, Kaluža B (2009) Fall detection and activity recognition with machine learning. Informatica 33(2):205–212
422. Kangas M, Vikman I, Wiklander J, Lindgren P, Nyberg L, Jämsä T (2009) Sensitivity and specificity of fall detection in people aged 40 years and over. Gait Posture 29(4):571–574
423. Sposaro F, Tyson G (2009) ifall: an android application for fall monitoring and response. In: 2009 Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 6119–6122
424. Nyan M, Tay FE, Murugasu E (2008) A wearable system for pre-impact fall detection. J Biomech 41(16):3475–3481
425. Bourke AK, O'Donovan KJ, Olaighin G (2008) The identification of vertical velocity profiles using an inertial sensor to investigate pre-impact detection of falls. Med Eng Phys 30(7):937–946
426. Kangas M, Konttila A, Lindgren P, Winblad I, Jämsä T (2008) Comparison of low-complexity fall detection algorithms for body attached accelerometers. Gait Posture 28(2):285–291
427. Wu G, Xue S (2008) Portable preimpact fall detector with inertial sensors. IEEE Trans Neural Syst Rehabil Eng 16(2):178–183
428. Bourke AK, O'brien J, Lyons GM (2007) Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait Posture 26(2):194–199
429. Jafari R, Li W, Bajcsy R, Glaser S, Sastry S (2007) Physical activity monitoring for assisted living at home. In: 4th International workshop on wearable and implantable body sensor networks (BSN 2007). IEEE, pp 213–219
430. Kangas M, Konttila A, Winblad I, Jamsa T (2007) Determination of simple thresholds for accelerometry-based parameters for fall detection. In: 2007 29th Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp 1367–1370
431. Lindemann U, Hock A, Stuber M, Keck W, Becker C (2005) Evaluation of a fall detector based on accelerometers: a pilot study. Med Biol Eng Comput 43(5):548–551 (**Springer**)
432. Chen J, Kwong K, Chang D, Luk J, Bajcsy R (2005) Wearable sensors for reliable fall detection. In: 2005 IEEE engineering in medicine and biology 27th annual conference. IEEE, pp 3551–3554
433. Zhuo S, Sherlock L, Dobbie G, Koh YS, Russello G, Lottridge D (2020) Real-time smartphone activity classification using inertial sensors-recognition of scrolling, typing, and watching videos while sitting or walking. Sensors 20(3):655
434. Wen H, Ramos Rojas J, Dey AK (2016) Serendipity: finger gesture recognition using an off-the-shelf smartwatch. In: Proceedings of the 2016 CHI conference on human factors in computing systems. ACM, pp 3847–3851
435. Georgi M, Amma C, Schultz T (2015) Recognizing hand and finger gestures with IMU based motion and EMG based muscle activity sensing. In: Biosignals. Citeseer, pp 99–108
436. Bulling A, Blanke U, Schiele B (2014) A tutorial on human activity recognition using body-worn inertial sensors. ACM Comput Surv (CSUR) 46(3):1–33
437. Junker H, Amft O, Lukowicz P, Tröster G (2008) Gesture spotting with body-worn inertial sensors to detect user activities. Pattern Recognit 41(6):2010–2024
438. Chen C, Kehtarnavaz N, Jafari R (2014) A medication adherence monitoring system for pill bottles based on a wearable inertial sensor. In: 2014 36th Annual international conference of the IEEE engineering in medicine and biology society. IEEE, pp. 4983–4986
439. Bächlin M, Roggen D, Tröster G, Plotnik M, Inbar N, Maidan I, Herman T, Brozgol M, Shaviv E, Giladi N et al (2009) Potentials of enhanced context awareness in wearable assistants for Parkinson's disease patients with the freezing of gait syndrome. In: International symposium on wearable computers (ISWC). IEEE, pp 123–130
440. Jovanov E, Milenkovic A, Otto C, De Groen PC (2005) A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation. J NeuroEng Rehabil 2(1):1–10