Application of Smart Insoles for Recognition of Activities of Daily Living: A Systematic Review

D'Arco, L., McCalmont, G., Wang, H., & Zheng, H. (2023). Application of Smart Insoles for Recognition of Activities of Daily Living: A Systematic Review. *ACM Transactions on Computing for Healthcare*, 1-37. Advance online publication. https://doi.org/10.1145/3633785

Published in:
ACM Transactions on Computing for Healthcare

Publication Status:
Published online: 24/11/2023

DOI:
10.1145/3633785

Document Version
Author Accepted version

Application of Smart Insoles for Recognition of Activities of Daily Living: A Systematic Review

LUIGI D'ARCO, Ulster University, UK
GRAHAM MCCALMONT, Ulster University, UK
HAIYING WANG, Ulster University, UK
HUIRU ZHENG*, Ulster University, UK

Recent years have witnessed the increasing literature on using smart insoles in health and well-being, and yet, their capability of daily living activity recognition has not been reviewed. This paper addressed this need and provided a systematic review of smart insole-based systems in the recognition of Activities of Daily Living (ADLs). The review followed the PRISMA guidelines, assessing the sensing elements used, the participants involved, the activities recognised, and the algorithms employed. The findings demonstrate the feasibility of using smart insoles for recognising ADLs, showing their high performance in recognising ambulation and physical activities involving the lower body, ranging from 70% to 99.8% of Accuracy, with 13 studies over 95%. The preferred solutions have been those including machine learning. A lack of existing publicly available datasets has been identified, and the majority of the studies were conducted in controlled environments. Furthermore, no studies assessed the impact of different sampling frequencies during data collection, and a trade-off between comfort and performance has been identified between the solutions. In conclusion, real-life applications were investigated showing the benefits of smart insoles over other solutions and placing more emphasis on the capabilities of smart insoles.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Human-centered computing → Ubiquitous and mobile computing**; • **Computing methodologies → Machine learning algorithms**.

Additional Key Words and Phrases: Systematic Review, Human Activity Recognition, Smart Insoles, Machine Learning

1 INTRODUCTION

Computers today are within everyone's reach, and the way we interact with them is evolving. A computer can be defined as anything that is able to perform calculations, manipulate or process data, and control continuous or discrete processes [19]. The goal is for computers to fade into the background of our daily lives, becoming a part of our environment [87]. Reducing the number of explicit interactions required for the user to communicate with the computer while increasing the number of implicit interactions is one technique for achieving this goal. An ideal system should be able to recognise the user's demands by analysing the user's current state and surroundings. As a result, the researchers' interest throughout the years has focused on recognising human activities in controlled and uncontrolled environments. The objective of Human Activity Recognition (HAR) is the creation of predictive models that allow the classification of the behaviour of individuals, by means of a behavioural model such as the one proposed by Fogg [28], making it possible to identify the time when a

*Corresponding author

Authors' addresses: Luigi D'Arco, darco-l@ulster.ac.uk, Ulster University, York St., Belfast, Northern Ireland, UK, BT15 1ED; Graham McCalmont, mccalmont-g@ulster.ac.uk, Ulster University, York St., Belfast, Northern Ireland, UK, BT15 1ED; Haiying Wang, hy.wang@ulster.ac.uk, Ulster University, York St., Belfast, Northern Ireland, UK, BT15 1ED; Huiru Zheng, h.zheng@ulster.ac.uk, Ulster University, York St., Belfast, Northern Ireland, UK, BT15 1ED.

ACM 2637-8051/2023/11-ART
https://doi.org/10.1145/3633785

certain action is being taken by the user and therefore to decide when intervention to assist him is necessary. The approaches used can be classified into two main categories in terms of the devices used, i.e. vision-based HAR and sensor-based HAR [15]. Vision-based HAR systems analyse images or videos obtained from capture devices [9, 53], whereas sensors-based HAR systems focus on processing data extracted from wearable/environmental sensors [25, 70]. The former systems comprise, in turn, different systems that can be differentiated based on the type of sensors used and the relative data produced: 2D images [79], 3D images [8], infrared images [89], or videos [66]. The latter systems, instead, are mainly based on the type of sensors involved: wearable sensors [29], sensors on objects [38], or ambient sensors [6]. Initially, the most adopted solution was the vision one as it reached higher performances and it was easier to analyse the results. However, they are significantly affected by challenges such as occlusion, anthropometry, execution rate, background clutter, and camera motion [11], and they could create privacy and acceptability problems for users [78]. Given the hardware improvements of the sensors, the reduction in their sizes and costs and the increase in their processing capacity, sensor-based systems have become a viable solution that represents a minimally invasive and easier acceptance solution for the user, but also a solution more adaptable to different environments.

Although HAR is used in many fields, including fitness [58], home automation [74], and security [49], the most sought-after application in which to implement HAR is the integration into daily life to monitor and evaluate the activities that a user carries out. This activity can be evaluated either for prevention, such as fall detection [43, 55], or for monitoring, expenditure estimation [4, 20], stress detection [30], behaviour monitoring [68], or for rehabilitation [1].

The most recent HAR solutions aim to use wearable sensors, unlike environmental sensors and cameras, as they can be installed regardless of the environment in which they will be used, thus providing more degrees of freedom on possible applications. Wearable sensors can be placed at different points of the user's body, including the head, chest, wrist, waist and ankle [48]. Wearable sensors differ from each other not only by their position but also by the type of sensors used. The choice of the sensors plays an important role in activity recognition performances [77]. The popular types are the Inertial Measurement Units (IMUs), which when installed in a device, capture data about the user's movement. IMUs usually contain sensors such as accelerometers, gyroscopes, and magnetometers [26], that respectively measure the rate of change of the velocity, the three-axial angular velocity and the change of the magnetic field.

Determining what the least invasive position is and which allows obtaining a greater yield of information is a very coveted goal that has been extensively studied over time. In [14], Cleland et al. analysed what are the optimal positions for using an accelerometer, as well as in [92] Faruk et al. analysed the minimum number of inertia sensors required for activity recognition. Using multiple devices allows for increasing the accuracy of the results, however, creates impediments in the user's movement, which makes it impossible to use in everyday life due to the inconvenience they can cause.

In this regard, a technology that has been emerging in recent years is the smart insole, as it allows the integration of multiple sensors, limiting the occurrence of impediments for the user. Smart insoles can generally be defined as high-tech inserts that can be placed inside any shoe and are designed to monitor and collect data related to movements and pressure distribution. The advantages of this solution include the high number of sensors that can be used without affecting the user's mobility, the simplicity of installation in an uncontrolled environment, and greater acceptance by the user since once installed there are no more operations to do except for recharging it [50].

Smart insoles were first introduced in 2001 in the study conducted by Pappas et al. [67] which aimed to detect in real-time the gait phases including stance, heel-off, swing and heel-strike by using three force sensors resistor and a gyroscope. Further work was released later in 2005 by Zhang et al. [90] in which an artificial neural network was applied for the identification of the type and intensity of locomotion, but only since 2008, the interest in them has grown exponentially thanks to the introduction of flexible materials and the miniaturisation of electronic devices [7, 45]. With the maturity of the technology, smart insoles have reached levels of accuracy and precision comparable to much larger devices mainly used in the biomechanical fields, as demonstrated in 2021 by Guo et al. [32], who compared the performance of smart insoles with force plates.

Smart insoles are primarily applied to recognise Activities of Daily Living (ADLs), which refer to routine tasks that individuals perform daily to maintain their self-care and independence [60]. These activities cover a wide variety of tasks like dressing, feeding, and bathing as well as ambulation activities (commonly referred to as functional mobility), such as walking, running, going up and down stairs, standing, sitting, etc. Although the use of smart insoles for the recognition of ADLs is the topic of this review, it should be considered that the applications of these devices extend beyond ADLs recognition. They can be used to detect falls [51], a crucial application for the elderly population or people with neurological disorders, where early detection can lead to timely medical intervention [73]. Smart insoles can also be used for the treatment and recognition of gait-related diseases, such as Parkinson's disease or multiple sclerosis, by providing real-time feedback on an individual's gait and balance [18]. Additionally, they have the potential to be used in sports and fitness settings, where they can provide insights into an athlete's performance [86].

The literature includes a variety of reviews analysing the effectiveness of smart insoles in multiple contexts. Almuteb et al. [2] summarised the prototypes and the commercial solutions that are referred to as smart insole in literature. Despite the review having covered multiple areas, focusing more on the hardware areas and touching on the possible application of such smart insoles, including human activity recognition, the manuscript lacks an in-depth discussion on the settings and the algorithms used in each article. Ngueleu et al. [63] presented a systematic review of instrumented insoles. They analysed the solutions proposed in 33 papers, but that spanned between multiple arguments, including step counting, posture and activity recognition. Although they analysed the various sensors included in the relative studies, they reviewed the recognition of various activities including but not limited to daily activities making it difficult to assess the validity of smart insoles in daily living. This review was carried out in 2019, and covered only papers until the first quarter of the same year, leaving outside all the papers that have been published since then. Furthermore, we believe that the papers published before 2015 are of less interest, given the rapid development of the topic and the improvements in electronic devices. Subramaniam et al. [82] provided a detailed review of insole-based systems applied in the monitoring of plantar pressure, activity, and gait. They analysed the research gaps and challenges of the topic, but they covered mostly multi-modal systems in which smart insoles have been included.

Although all the state-of-the-art reviews analysed are of great value and adequately address the chosen topics, there is a lack in the literature of a detailed analysis of what are the capabilities of smart insoles in everyday use when integrated as a single device for the recognition of daily living activities.

The aim of this systematic review was to investigate the capability of smart insoles to recognise ADLs by comparing and analysing existing solutions in the literature in terms of sensors comprising the smart insole, settings chosen, algorithms and performance reached. The peculiarity of this review, besides the type of activities included, is the evaluation of smart insole as the only device for recognition, to provide an effective analysis of their capability in real-life applications. Hence, the following research questions were developed to guide the review:

RQ 1. Is it feasible to use smart insoles in human activity recognition?
RQ 2. Which types of activities of daily living can be recognised by using only smart insoles?
RQ 3. What are the limitations of current solutions and what are the drawbacks of the currently used algorithms?

The systematic review was carried out using the PRISMA guidelines [65], including only studies in which the smart insoles are used as the only device and in which the definition of ADLs was met. Thus any solution that includes not only smart insoles but any other type of device except for the ankle is discarded if the performance of only smart insoles cannot be extracted. Similarly, articles that do not address or individually examine ADLs were discarded.

From the findings of the review, it was possible to discuss the choices made in existing studies, highlighting the challenges and gaps that have yet to be addressed, as well as the scenarios in which smart insoles can have a significant impact, and providing readers suggestions to help them speed up their future research.

This systematic review is organised as follows. Section 2 describes the methodology used for the discovery, filtering and evaluation of existing solutions, Section 3 summarises the results of the review process, followed by Section 4 in which the main findings, challenges and gaps are presented. Section 5 concludes the paper.

2 METHODOLOGY

2.1 Data Source and Search Strategy

This Systematic Review has been carried out following the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [65]. The articles analysed were extracted from five different databases: Scopus, Web of Science, PubMed, Compendex and Embase. The following terms were used to search the databases: ("activity detection" OR "activity recognition" OR "activity classification" OR "activity identification") AND (insole* or shoe* or "plantar pressure"). The search strategy included all the papers that had been published before 31 October 2022.

2.2 Inclusion/Exclusion Criteria

Studies were included when: (1) smart insoles were used for activity recognition; (2) the activities considered were Activities of Daily Living (ADLs); (3) they were written in English; (4) they were published after 2015.

Studies that include multiple devices in addition to smart insoles, were considered only if the results from each device were distinguishable and the smart insoles results could be extracted. Similarly, for studies in which multiple activities including ADLs were classified, only those in which the results in ADLs were well-defined and could be extracted were included in this review.

Articles were excluded if (1) they were published before 2015; (2) they were usability studies; (3) they involved multiple devices besides smart insoles and the results were not well separated; (4) they were written in a language different from English (5) they didn't involve ADLs or ADLs were not well defined (6) they were conference paper and the extended paper has already been included.

2.3 Data Extraction and Selection

The articles resulting from the database search strategy were analysed and the duplicates were removed. Two independent researchers, reviewers 1 and 2, evaluated the titles and the abstracts of the remaining articles and based on the inclusion criteria they screened the articles. The full text of the remaining articles was retrieved and assessed for eligibility by the reviewers. Finally, to consolidate the work done, the two reviewers discussed the discrepancies in the selected articles, defining a new list of articles accepted for the study. Any doubts and uncertainties regarding the inclusion of a particular study were discussed with reviewers 3 and 4. The relevant data were extracted from the accepted articles following a standardised form. The data included article reference, sensing elements involved with technical specification (e.g., sample rate, transmission method), participants' characteristics, activity types, algorithms used, validation strategy, and recognition performance (e.g., accuracy, precision, specificity, sensibility and F1-score).

Table 1. Quality Assessment Items

Number	Item
1	Is the question/objective sufficiently described?
2	Is the study design evident and appropriate?
3	Are the subjects' (and comparison group, if applicable) characteristics sufficiently described?
4	Is the nature of the data used for recognition and how they were determined well described?
5	Are the devices involved well presented?
6	Is the choice of the algorithm well explained?
7	If the algorithm used is bespoke, is it compared to off-the-shelf algorithms?
8	Are outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias? Are the means of assessment reported?
9	Is the sample size appropriate?
10	Are the analytic methods described/justified and appropriate?
11	Are some estimates of variance reported for the main results?
12	Are the findings controlled for confounding?
13	Are the results reported in sufficient detail?
14	Are the conclusions supported by the results?

2.4 Quality Assessment

The inclusion of a variety of databases has guaranteed a more thorough search, achieving greater levels of sensitivity, and reducing source publication bias. A quality assessment method was created to score and rank the accepted studies in order to minimise biased judgement.

The quality assessment method used is adapted from the quality appraisal tool proposed by Kmet et al. [44]. It assesses the methodological quality of quantitative and qualitative articles as well as the likelihood of bias. The scoring system is composed of 14 quality assessment items (Table 1), and depending on the degree to which the specific criteria were met, the score can be selected from the following: 2 - completely satisfied; 1 - partially satisfied; 0 - not satisfied. If an item is not addressed within the study, the N/A score is assigned with a value equal to 0 in the calculation of the final score.

For each article, the quality score ($Q.Score$) was calculated as expressed in percentage using the following equation:

$$Q.Score = \frac{\sum_{i=1}^{n} s_i}{max\_score} \tag{1}$$

where s_i is the score for the i-th item, the max_score is the maximum score that a study can obtain from the quality assessment (28 for 14 quality assessment items in this review), and n is the number of quality assessment items (14 in this review).

Moreover, qualitative categories are identified to improve the interpretation of the articles analysed based on obtained quality values [21]. The study quality was calculated as follows:

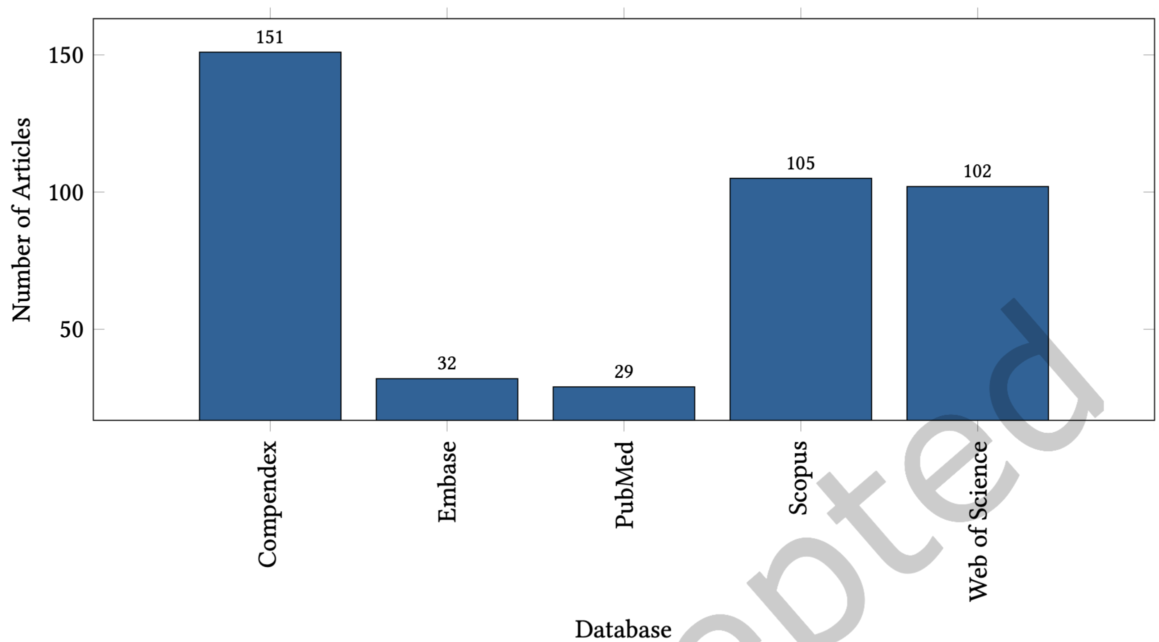


Fig. 1. Number of articles for each database identified from the search strategy.

$$StudyQuality = \begin{cases} \text{High Quality (HQ)} & \text{if } Q.Score \geq 95\% \\ \text{Good Quality (GQ)} & \text{if } 85\% \leq Q.Score < 95\% \\ \text{Moderate Quality (MQ)} & \text{if } 65\% \leq Q.Score < 85\% \\ \text{Low Quality (LQ)} & \text{if } Q.Score < 65\% \end{cases} \quad (2)$$

3 RESULTS

The initial search strategy resulted in 419 articles from the selected databases. The number of articles obtained for each database is shown in Fig. 1.

After removing the duplicate publications, the number of papers included in the review was 133, which were screened by title, abstract and publication year. An article was defined as *discarded* if any one inclusion criteria were not met, *accepted* if all the inclusion criteria were met, and *dubious* if the title and the abstract didn't provide sufficient information to decide if the article could be discarded or accepted. The number of articles remaining after this first screening was 61. For all these articles the full text was retrieved for further screening. The number of articles that met all the inclusion criteria was 26. The entire PRISMA process of searching, screening and selection is presented in Fig. 2.

3.1 Sensing Elements

The main goal of this review is to prove the importance of smart insoles for recognising ADLs, hence all the selected articles include a smart insole in the study. A smart insoles system is a system that has an insole as its main component where multiple sensors are housed inside. The sensors included vary from solution to solution,

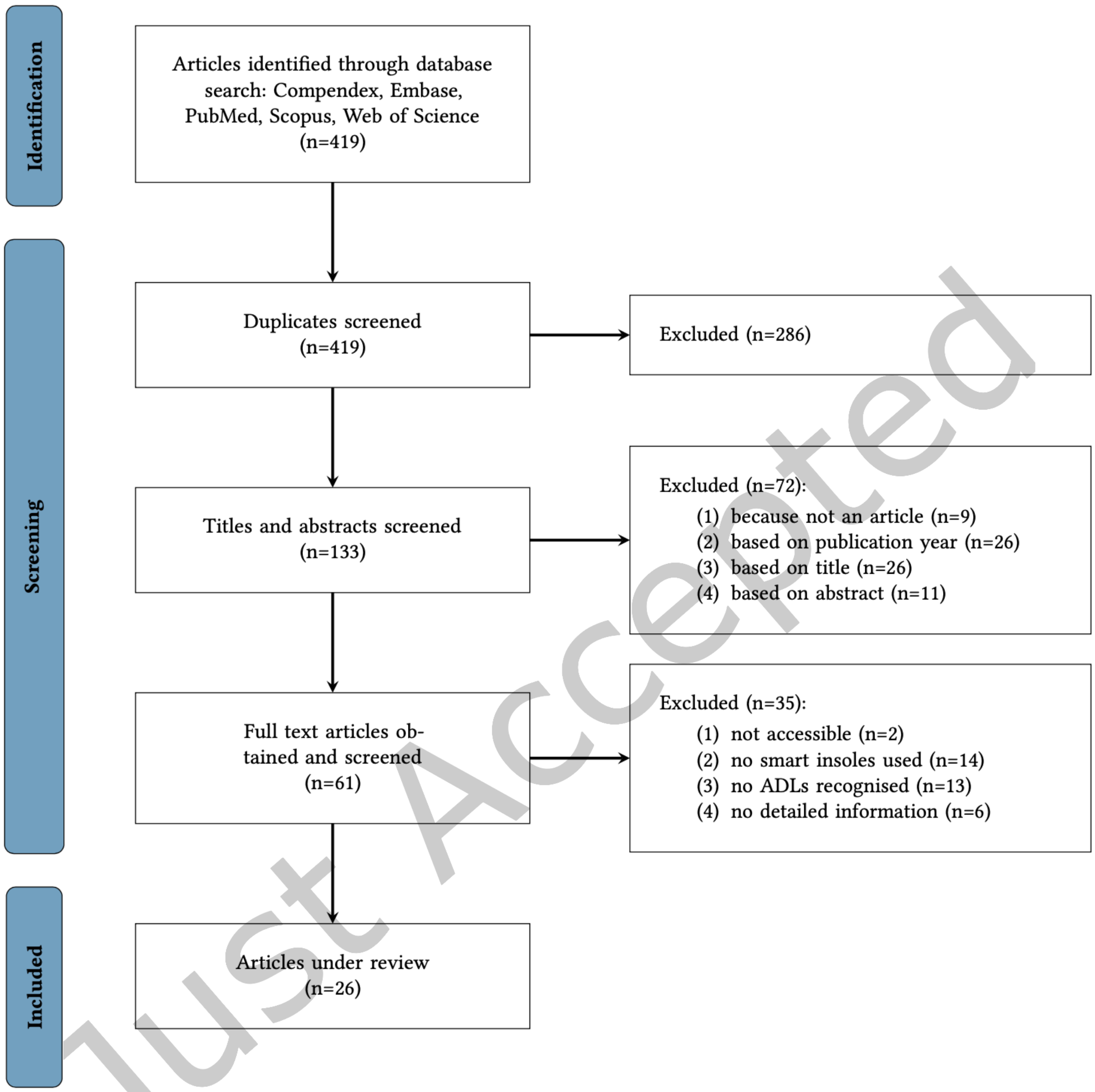
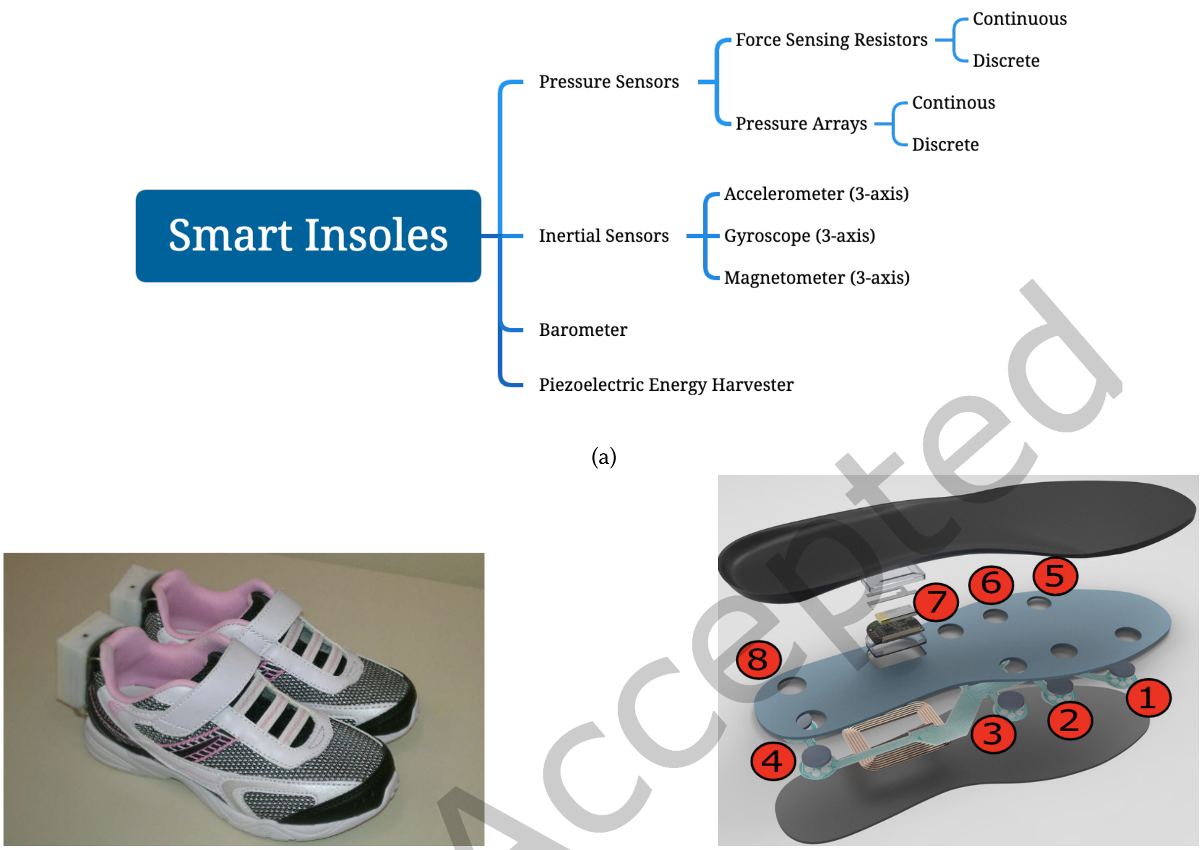


Fig. 2. Flowchart depicting the process of identifying and selecting articles following the PRISMA guidelines.

but mainly pressure sensors and inertia sensors can be observed. Although the purpose is to integrate everything necessary within the insole itself, several solutions have an additional section in which the electronic components, batteries and sometimes even sensors, such as inertia, are inserted. These additional devices are usually attached to the shoe or ankle to limit the footprint for the user.



(a)

(b)

(c)

Fig. 3. Illustration of smart insoles systems (a) sensors identified from the analysed studies (b) example of a smart insoles system with additional electronic component [37] (c) example of a smart insoles system with electronic encapsulated [64].

Among the 26 studies, eight of them preferred to use only pressure sensors, two of them only used IMU sensors and 14 of them used both types of sensors. Only one study used a completely different sensing element [47], which was based on energy harvesting and capacitor charging. The sensors identified in the analysed study are reported in Fig. 3a. Table 2 summarises the sensing elements and their settings from the 26 studies. The data include the type and number of sensors included, the sampling frequency and the data transmission technology involved in data collection. The pressure sensors used can be classified into two categories: force-sensing resistors (also known as pressure sensors) and pressure arrays, by the way the force is measured. The force-sensing resistors are placed on an insole in a specific part of the foot plantar as individual sensors, for example, Heel, Metatarsals, and Hallux. The pressure array is spread on a whole insole surface and it can measure the force from the entire foot. Among those analysed, only three studies [12, 13, 59] involved pressure arrays, however in the study presented by Merry et al. [59] the pressure array was used to create 10 anatomical regions of interest to extract and simulate 24 force-sensing resistors. Instead, the 16 studies that included pressure sensors differ from each other based on the location and the number of pressure sensors used, ranging from two [40] to 21 pressure sensors [80]. The

Table 2. Summary of the sensing elements and their settings involved in the included studies.

Study	Sensing Elements		Sampling Frequency (Hz)	Data Transmission
	Pressure Sensors	**Inertia Sensors**		
Chen et al. [13]	PSA (up 96 PS)	3D-ACC, 3D-GYR	30	Bluetooth
Chen et al. [12]	PSA (up 96 PS)	3D-ACC, 3D-GYR	30	Bluetooth
D'Arco et al. [17]	8 PS (Hallux, Toes, 1°/3°/5° metatarsal, arch, lateral and medial heel)	3D-ACC, 3D-GYR, 3D-MAG	200	BLE
De Pinho et al. [5]	6 PS	3D-ACC, 3D-GYR, 3D-MAG, BAR	10	Wi-Fi
Dehzangi et al. [22]	13 PS	3D-ACC	50	ANT+, Flash memory
Gonzalez et al. [31]	4 PS	3D-ACC	50	Bluetooth
Haescher et al. [34]	6 PS	/	20	USB
Hedge et al. [36]	3 PS (Heel, Metatarsal Head, Big Toe)	3D-ACC	50	BLE
Hedge et al. [37]	5 PS (Heel, 1°/3°/5° metatarsal, Big Toe)	3D-ACC	400, downsampled to 25 averaging 16 consecutive samples	Bluetooth
Hegde et al. [35]	3 PS (Heel, 1° Metatarsal Head, Big Toe)	3D-ACC	50	BLE, Flash memory
Jeong et al. [39]	8 PS	/	50	/
Key et al. [40]	2 PS	/	1000	Wi-Fi
Lan et al. [47]	2 PEH Capacitors		100	Flash Memory
McCalmont et al. [57]	8 PS (Hallux, Toes, 1°/3°/5° metatarsal, arch, lateral and medial heel)	3D-ACC, 3D-GYR, 3D-MAG	/	BLE
Merry et al. [59]	PSA (10 anatomical regions of interest)	/	75 with a reduction to 15 averaging 5 consecutive frames	/
Moufawad el Achkar et al. [61]	8 PS	3D-ACC, 3D-GYR, 3D-MAG, BAR	200	/
Moufawad el Achkar et al. [62]	8 PS (Hallux, Toes, 1°/3°/5° metatarsal, arch, lateral and medial heel)	3D-ACC, 3D-GYR, 3D-MAG, BAR	200	Flash Memory
Nguyen et al. [64]	8 PS	/	50	Bluetooth
Paydafar et al. [69]	3 PS (calcaneus, metatarsal and phalanges)	/	50	Wireless
Ren et al. [72]	7 PS	/	100	Wireless
Sazonov et al. [75]	5 PS	3D-ACC	400, downsampled to 25 averaging 16 consecutive samples	BLE
Song et al. [80]	21 PS	/	/	/
Truong et al. [83]	8 PS	3D-ACC, 3D-GYR	50	Bluetooth
Truong et al. [84]	8 PS	/	50	Bluetooth
Wang et al. [85]	/	3D-ACC, 3D-GYR	50	BLE
Zhang et al. [91]	/	3D-ACC, 3D-GYR	100	/

3D-ACC: 3-axis accelerometer; 3D-GYR: 3-axis gyroscope; 3D-MAG: 3-axis magnetometer; BAR: barometer sensors; BLE: Bluetooth low energy; PEH: piezoelectric energy harvester; PS: pressure sensor; PSA: pressure sensor array.

only exception is the study proposed by Truong et al. [84] in which the pressure sensors provided values in four discrete intensive levels, between 0 and 3, instead of providing continuous values.

The subject's movement can be measured using IMU sensors. The accelerometer and gyroscope, which measure acceleration in terms of spatial and angular velocity, are the most commonly employed sensors in the studies evaluated, with the accelerometer being used exclusively in six studies and the accelerometer and gyroscope being used together in five studies. In more complex systems, a magnetometer is added to the accelerometer and gyroscope to improve the information generated by taking into account magnetic field measurements [17, 57], or additionally, a barometer is added to measure altitude variation [5, 61, 62].

Lan et al. [47] opted for a completely different sensing architecture. They proposed the use of two piezoelectric energy harvesting (PEH) transducers, which harvest energy from the ground reaction pressure associated with foot movement. By calculating the voltage increment rate in the capacitors generated by the PEH transducers, they can recognise the activity carried out by the user.

Even if a system based on smart insoles is present in all the studies, the configurations and locations of the sensors and electronic components vary. The most frequent scenario is when the electronic component is tethered to the ankle or fastened to the shoe (on the side, top, or back), as shown in Fig 3b. However, this approach limits the user's mobility. Although they have limited performance, a number of solutions have been put forth where everything, from pressure sensors to inertia sensors passing via the control unit and the battery, is contained inside one insole, as shown in Fig. 3c. Among the various solutions, it is worth noting the one proposed by Wang et al. [85], which integrated only inertia sensors into the insole, such as an accelerometer and gyroscope, and the solution proposed by Chen et al. [13], which not only integrate a PSA with 96 pressure sensors but also included an accelerometer and a gyroscope all encapsulated inside an insole.

The sampling of the data is regulated by a frequency value, the higher the frequency, the more energy consumption will be, but more data are gathered. The sampling frequency of the experiments examined ranges from 10 [5] to 1000 Hz [40]. The most commonly utilised sample frequency is 50 Hz, which has been employed in 10 studies. Three research employed a sampling frequency of 100 Hz [47, 72, 91], and further three used one of 200 Hz [17, 61, 62]. Two studies used a sampling frequency of 30 Hz [12, 13], and two studies used 400 Hz [37, 75]. The rest opted for different sampling frequencies, Merry et al. used the 75 Hz [59], Haescher et al. used the 20 Hz [34] and De Pinho et al. used the 10 Hz [5]. Only two studies [57, 80] didn't specify the sampling frequency. In addition, three studies applied signal averaging to increase the strength of a signal relative to noise that is obscuring it. Sazonov et al. [75] and Hedge et al. [37] down-sampled from 400 Hz to 25 Hz by averaging 16 consecutive samples, whereas, Merry et al. [59] down-sampled from 75 Hz to 15 Hz by averaging five consecutive samples.

Once the data have been collected, they need to be processed. Since an insole-based system has no sufficient computational power to support that, the data are stored or sent to an external device such as a computer or a smartphone. The most broadly utilised transmission technology is Bluetooth, which was employed in 13 studies, including six studies that used the low-energy alternative, Bluetooth Low Energy (BLE). The remaining analysed studies employed instead disparate technologies. Two studies used Wi-Fi [5, 40]. One study used ANT+ [22], which is an ultra-low power transmission technology developed specifically for the health, fitness and sports segment [42]. Four studies opted to store the data on a flash memory offline [22, 35, 47, 62]. One study involved a USB connection for directly storing data on the computer [34]. Two studies defined that the data were transmitted wirelessly to an external device but without defining the exact technology [69, 72], whereas, five studies didn't include the transmission technology in their articles.

Table 3. Summary of the number of participants with relative demographic information involved in the included studies.

Study	No. Participants	Gender	Typology	Participants Demography				
				Age	**Height**	**Weight**	**BMI**	**Shoe Size**
Chen et al. [13]	10	/	Healthy	/	/	/	/	US 5.5 - 14
Chen et al. [12]	10	/	Healthy	/	/	/	/	US 5.5 - 14
D'Arco et al. [17]	5	/	Healthy	25-55	/	/	/	/
De Pinho et al. [5]	11	/	5 Healthy, 2 Elders, 1 Obese, 3 Knee-injured	/	/	/	/	US 7.5 - 8.5
Dehzangi et al. [22]	10	/	Healthy	/	/	/	/	/
González et al. [31]	5	5 M	Healthy	33 ± 2	/	/	/	EU 41-44
Haescher et al. [34]	13	12 M, 1 F	Healthy	22 - 49	172 - 192	58-93	/	EU 43
Hedge et al. [36]	4	3 M, 1 F	Healthy	28 ± 0.5	170 ± 4	69.2 ± 12.7	24.2 ± 3.7	US M9, US W9
Hedge et al. [37]	21	12 M, 9 F	11 Healthy, 10 Cerebral Palsy (CP)	Healthy: 6.6 ± 1.5; CP: 6.2 ± 1.5	Healthy: 120 ± 10; CP: 120 ± 10	Healthy: 24.4 ± 4.4; CP: 22.3 ± 4.6	/	US kids 12 - Youth 2
Hegde et al [35]	15	8 M, 7 F	Healthy	M: 26.6 ± 3.4, F: 23.3 ± 5	M: 180 ± 5, F: 165 ± 8	M: 81.9 ± 17.2, F: 66.7 ± 9.9	M: 21.9 ± 4.5, F: 24.7 ± 5.4	M: US M8 - M11, F: US W6 - W9
Jeong et al. [39]	3	/	Healthy	/	/	/	/	/
Key et al. [40]	1	/	Healthy	/	/	/	/	/
Lan et al. [47]	10	8M, 2 F	Healthy	24 -30	168 - 183	55 - 75	/	/
McCalmont et al. [57]	1	1 M	Healthy	/	/	/	/	/
Merry et al. [59]	8	7 M, 1 F	Healthy	29.0 ± 4.8	172.0 ± 10.3	74.2 ± 18.4	/	/
Moufawad el Achkar et al. [61]	10	8 M, 2 F	Elderly	65 - 75	162 - 184	62 - 114	/	/
Moufawad el Achkar et al. [62]	10	8 M, 2 F	Elderly	69.9 ± 3.1	171,7 ± 8.9	80.1 ± 14,7	/	EU 39 - 45
Nguyen et al. [64]	3	3 M	Healthy	24 - 29	167.0 ± 5	67.0 ± 9	/	/
Paydarfar et al. [69]	20	/	Healthy	20 -35	/	/	/	/
Ren et al. [72]	17	17 F	Healthy	26 ± 9	/	49 ± 3	/	22 - 27 cm
Sazonov et al. [75]	19	10 M, 9 F	Healthy	M: 28.1 ± 6.9, F: 23.7 ± 3.1	M: 178.1 ± 10.2, F: 167.5 ± 8.5	M: 79.3 ± 16.7, F: 72.0 ± 18.0	M: 24.8 ± 3.6, F: 25.6 ± 6.3	7 US W - 12 US M
Song et al. [80] 30	/	Healthy	23 - 40	M: 169 ± 3.74, F: 159 ± 2.45	M: 67.7 ± 3.88, F: 52 ± 1.93	/	/	EU M: 41.5, F: 36
Truong et al. [83]	2	/	Healthy	/	/	/	/	/
Truong et al. [84]	29	23 M, 6 F	Healthy	/	/	/	/	/
Wang et al. [85]	10	5 M, 5 F	Healthy	M: 22 - 40, F: 22 - 40	M: 170 - 185, F: 158 - 175	M: 58 - 90, F: 47 - 55	/	/

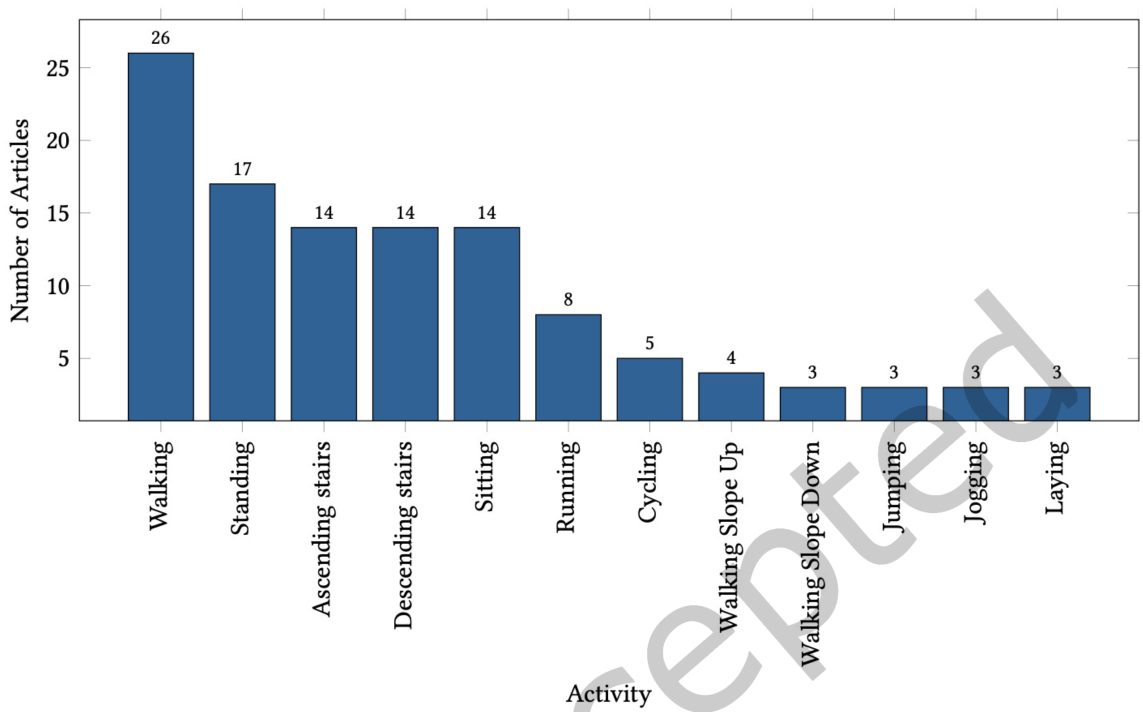


Fig. 4. Most recognised activities in the selected articles.

Zhang et al. [91]	8	8 M	Healthy	24.13 ± 3.47	176.38 ± 4.92	72 ± 10.87	/	/

F: female; M: male.

3.2 Participants

Understanding which group of subjects is involved is very important to determine the effectiveness of a study. If the participants are mostly heterogeneous, this can lead to an increase in the robustness of the proposed system. Unfortunately, not all the selected articles reported the demography of participants, only 22 studies provided the information.

As shown in Table 3, 14 studies recruited adults aged between 19 and 49, one study included only children aged between 5 and 7 [37], and three studies involved participants aged between 65 and 75 [5, 61, 62]. All the studies involved healthy subjects in their research; De Pinho et al. [5] had participants comprising healthy people, elderly adults, obese people and knee-injured subjects, while, Hegde et al. [37], involved participants with cerebral palsy.

The average number of participants involved in all the studies was 10, ranging from one study having 30 participants [80], to studies including only one participant [40, 57].

Not all studies specify the gender information of the participants. Among those including gender information, only the study proposed by Ren et al. [72] was focused entirely on women, while in all other studies, there was a very high male ratio.

3.3 Activity Types

Because smart insoles are used to assess activities, we cannot expect the activities that are identified to be highly complicated, as there is no way to distinguish, for example, upper limb activities carried out while still standing.

Fig. 4 shows the top activities, that have been used for recognition in the selected articles. The most common one is the walking activity, which is presented in all the studies selected, followed by standing, ascending and descending stairs, and sitting.

It has been observed that the activities analysed in the selected articles are almost basic activities, except few articles that included more complex activities. Hedge et al. [35] included vacuuming, shelving items, washing dishes and sweeping the floor, whereas, Truong et al. [83] classified book loading, door opening, tooth brushing, mopping and window cleaning. Moreover, Wang et al. [85] presented a solution to distinguish between ADLs and falls, and Chen et al. [12] have included together with ADLs, hazard activities such as slip hazard and trip hazard

The activities used in the selected articles are summarised in the second column of Table 4.

3.4 Data Collection

Proper identification and definition of test sessions can improve and increase the robustness of the proposed solution. Although the number of participants differs from study to study, the types of tests performed are similar. Among the studies analysed there is a high rate of tests carried out in controlled environments, such as in laboratories or in specific environments where preventive measures have been taken. In all studies, the designated location for data collection is indoors, with the exception of the study proposed by Lan et al. [47] which used controlled environments, but in both indoor and outdoor sessions. In addition, three studies combined the collection of data in controlled environments with free living sessions for each participant [35, 61, 75], whereas, Zhang et al. [91] left full freedom to the users on how to complete the activities. The settings that each study adopted for data collection were reported in Table 4.

The methodologies by which the data were collected vary according to the study. The most common approach among the studies analysed was to have the same participant repeat the activities multiple times. The repetitions varied from a minimum of 2 times for activities lasting one minute [59] to a maximum of 20 times [84]. The distance covered during the walking/running sessions varies from 3 meters [57] to 30 metres [13], while for the activities of descending and ascending the stairs, the duration of the session was evaluated based on the number of steps on the staircase, which vary from 9 [13] to 17 [64]. The longest measurements identified were those in which the participant was asked to wear the system during a free-living environment, ranging from a minimum of one hour [75] to a maximum of a full day [35]. Additionally, in the study presented by Wang et al. [85], 800 trials of falls and 400 trials of ADLs were collected. The ways in which each study conducted its data collection are reported in Table 4.

Validating the collected data and labelling them is of fundamental importance to optimise and make the proposed system reliable. Among the studies analysed, the most common methodology for collecting data is to have each participant perform a specific activity and save the data with the relative label. This type of validation is mainly done manually by researchers. In six studies [17, 22, 36, 57, 64, 84], a smartphone application was adopted to automate the labelling process, allowing users to start the recording process and add labels to the record. In three studies [17, 36, 84], they preferred to let the participants themselves use the app, instead in [22, 57, 64] a supervisor of the study used the application. In only two cases [13, 59], the data collection sessions were recorded using a camera, and the labels for each sample were identified after analysing the videos. Furthermore, two articles [35, 83], carried out the data collection by using multiple devices including an activity tracker. The activity tracker can be considered a silver-standard device, which is a device that can provide moderately reliable information and can be used as ground truth for validating the data collected. In no study, a gold-standard device, a device that provides high precision information, such as the GAITRite mat as used in [3] has been involved.

Table 4. Summary of the data collection activities involved in the included studies

Study	Activities	Test Settings	Test Labelling	Test Methodology
Chen et al. [13]	Dynamic Activities: Walking, Running, Descend Stairs, Ascend Stairs	Controlled Environment	Video	Each subject performed the walking and running activities along a straight hallway of about 30m for three times. For descending and ascending stairs each subject performed each activity along a stair of nine steps 10 times
	Quasi-static Activities: Sitting, Standing			Each subject performed 5s of sitting, followed by 5s of standing, with stand-up and stand-down interleaving activities. The whole process has been repeated 20 times
Chen et al. [12]	Walking, Running, Descending Stairs, Ascending Stairs, Slip Hazard, Trip Hazard	Controlled Environment	Manually	Each subject performed the walking and running activities along a straight hallway of about 30m for three times. For descending and ascending stairs each subject performed each activity along a stair of nine steps 10 times. For walking on a slippery surface the subject walked along a straight slippery path of 4m long 10 times. The "slip hazard" was created by spraying detergent on a mosaic-tiled surface. Trip hazard activity was created by using a fixed box of 14 cm heights
D'Arco et al. [17]	Walking (slow, normal, fast), Ascending stairs, Descending stairs, Sitting to Standing, Sitting, Standing	Controlled Environment	App controlled by participant	Each participant chose the activities to carry out from the entire set. A total of 120 minutes of data were recorded
De Pinho et al. [5]	Walking straight, walking slope up, walking slope down, ascending stairs, descending stairs and sitting.	Controlled Environment	Manually	Each participant performed 5 cycles of 10-minute walking, 8 minute of regular walking and 2 of fast walking. Each participant performed 5 cycles of slope walking and sitting, 4 minutes walking slope up, 4 minutes walking slope down and 2 minutes sitting. Each participant performed in sequence 1 minute of ascending stairs, 1 minute of sitting, 1 minute of descending stairs and 1 minute of sitting, all repeated 5 times
Dehzangi et al. [22]	sitting, standing, walking, running, jumping, cycling	Controlled Environment	App controlled by supervisor	Each participant performed each activity for 1 minute
González et al. [31]	walking forward, walking backwards, lateral walking (walking left), lateral walking (walking right), turning left, turning right, sitting down and standing up	/	Manually	/
Haescher et al. [34]	sneaking, normal walking, fast walking, jogging, walking while carrying weight, cycling	Controlled Environment	Manually	The participant performed each walking activity on a treadmill and the cycling activity on an exercise bike
Hedge et al. [36]	sitting, standing, walking and cycling	Controlled Environment	App controlled by participant	The sitting and standing activities had variations during the data collection (sitting/standing with and without fidgeting)
Hedge et al. [37]	Sitting, Standing, Walking	Controlled Environment	Manually	Each participant performed each activity for 2 minutes on a walkway with a useful area of 61x366cm
Hegde et al [35]	laying, sitting, standing, walking, descend stairs, ascend stairs, vacuuming, shelving items, washing dishes, sweeping the floor, driving automatic shift car	Hybrid Environment	Manually	Participants performed activities in a random order for approximately 2 hours. Participants wore the device for the entire day in a free-living condition producing a total of 132 hours of data
Jeong et al. [39]	Level walking, Stair descent, Stair Ascent	Controlled Environment	Manually	/
Key et al. [40]	Standing, Walking, Running, Jumping, Playing basketball and Playing soccer	Controlled Environment	Manually	Each participant performed each activity for 5 minutes
Lan et al. [47]	walking, running, ascending stairs, descending stairs, stationary (sitting, standing)	Controlled Environment (indoor and outdoor)	Manually	Each participant completed at least two data collection sessions for both indoor and outdoor environments. For walking, running, and stationary, each session lasted at least 20 seconds, whereas, for ascending/descending stairs each session lasted from 6 to 10 seconds depending on the number of steps and the walking speed of the subject. In total 210 sessions of data were collected

McCalmont et al. [57]	slow walking, normal walking, fast walking, walk upstairs, walk downstairs	Controlled Environment	App controlled by supervisor	Each participant performed each exercise 10 times. For walking activities, the participants walked to a marker 3 meters in front of them, turned back and come to the starting point. For descending and ascending stairs twelve flights were used
Merry et al. [59]	sitting, standing, walking	Controlled Environment	Video	Each participant performed each activity twice for about 1 minute each. All the process has been repeated twice
Moufawad el Achkar et al. [61]	sitting, standing, level walking, upstairs, downstairs, uphill, downhill, elevator up, elevator down	Controlled Environment	Manually	Each participant followed a predefined track collecting data for 1 hour
Moufawad el Achkar et al. [62]	sitting, standing, walking	Hybrid Environment	Manually	Each participant performed a standing-still activity for 5 seconds followed by a level walking for 10 straight steps. Each participant collected 4 hours of free-living activities
Nguyen et al. [64]	level ground, incline descent, incline ascent, stair descent, and stair ascent walking	Controlled Environment	App controlled by supervisor	Each participant performed a level ground walking for 20 meters and 17 steps of stairs for both descending and ascending stairs. The inclined walking activities were performed on a 15-meter incline ground of approximately 11.5°
Paydarfar et al. [69]	walking, standing, balancing on the left foot, balancing on the right foot, toe-up, ascending stairs	Controlled Environment	Manually	Each activity was performed and recorded for 45 to 120 seconds.
Ren et al. [72]	sitting, standing, walking on a flat surface, walking upstairs, walking downstairs, walking up a slope, running, cycling, office work	Controlled Environment	Manually	Each participant performed each activity for approximately 4 minutes, except for walking on a flat surface and running which were approximately 8 minutes. The order in which the 9 activities were completed was randomly selected for each subject
Sazonov et al. [75]	sitting, standing, walking/logging, cycling	Hybrid Environment	Manually	Each participant after 30 minutes of equilibration, which is not used in the analysis, performed 20 minutes of laying, 20 minutes of sitting watching TV, 20 minutes of sitting working at the computer, 10 minutes of quiet standing, 10 minutes of active standing, and 6 activities from a group of 8 activities 10 minutes each. Each participant performed one hour of free-living data collection
Song et al. [80]	Walking, Running, Upstairs, Downstairs	Controlled Environment	Manually	Each participant performed each activity 8 times
Truong et al. [83]	Book Carrying, Door Opening, PC Using, Standing, Tooth Brushing, Mopping, Windows Cleaning, Walking, Jogging	Controlled Environment	Manually	Each participant performed each activity for 10 seconds
Truong et al. [84]	Level Walking, Stair descent, Stair ascent	Controlled Environment	App controlled by participant	Each participant performed each activity from 12 to 20 times
Wang et al. [85]	Falls: Forward-laying, Backward-laying, Leftward-laying, Rightward-laying; ADL: Laying on the bed, Bowing, Walking, Jogging, Laying down	Controlled Environment	Manually	A total of 800 falls have been collected divided into 200 trials for lying on the bed, bowing and lying down, respectively. A total of 200 continuous walking and jogging trails were collected
Zhang et al. [91]	standing, walking, running, laying, walking downstairs	Uncontrolled Environment	Manually	Each participant collected 5 minutes of each activity

3.5 Data Segmentation

There are usually certain differences between the data collected from several participants. First, the collected data do not all have the same length, and a sample of data can represent multiple sequential activities. The data segmentation allows the data collected to be split into smaller sections, also known as windows, to overcome these problems. This technique also enables the computational time consumption to be reduced by reducing the data complexity.

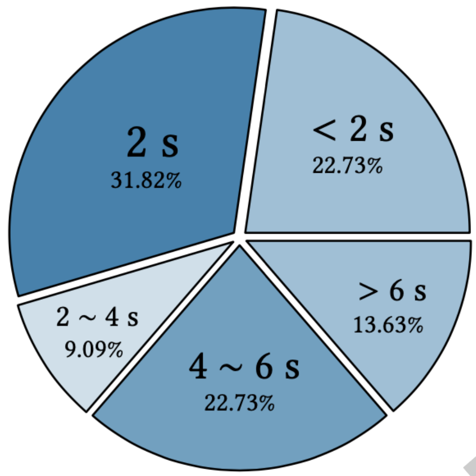


Fig. 5. Percentages of use of window sizes in studies where Time Windowing is used.

Multiple types of windowing techniques could be used, which differ from each other in the way the data are divided [71]:

- *Time Windowing* - the size of the segments is defined by an equal period of time;
- *Sensor Event Windowing* - the size of the segments is determined by a sequence of sensor events;
- *Dynamic Windowing* - the size of the segments could increase and decrease in length according to identified rules and threshold.

Twenty-three studies showed the use of a data segmentation technique while three studies [22, 57, 80] have not specified it. The data segmentation techniques used by each study have been reported in Table 6. The Time Windowing is definitely the favourite as 21 studies use it, followed by the Sensor Event Windowing which four studies used. Only in the study proposed by Cheng et al. [13] the two techniques have been used together. The Time Windowing technique for quasi-static activities, including standing and sitting, and the event Windowing technique for dynamic activities: including walking, running, ascending and descending stairs. Dynamic Windowing has not been applied in any study.

For the Sensor Event Windowing, the length of the window was based on the number of steps taken by the user, such as 1-6 steps in [39], 1-8 steps in [84], 6 steps in [64] and 30 steps in [12].

Concerning Time Windowing, there are no precise rules that allow establishing the optimal size of a window, this is because an activity is performed differently even by the same person over time [10]. The sizes varied in the studies because of the empirical choices that have been made. Fig. 5 shows the percentage of studies that use a specific window size. The size of the window varies according to the activity and the use that is made, ranging from 15 milliseconds [40] to 60 seconds [72]. However, the size of the window most adopted in the studies analysed is that of 2 seconds.

In data segmentation another parameter that must be considered is the percentage of overlap among consecutive windows, meaning that a percentage of the data in the previous window will be repeated in the next one. The overlap helps to eliminate the noise due to data truncation during window creation but also to increase the number of windows that are available for activity recognition. Among the studies analysed, three studies used

Table 5. Most used hand-crafted features in the included studies

Domain	Name	Formula	Description
Time	Mean	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$	It is the average of the collection of data
	Standard Deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$	It is a measurement of how much variance or dispersion there is in a set of data
	Variance	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	It is the average of squared deviations from the mean in a collection of data
	Skewness	$skw = \frac{1}{N\sigma^3}\sum_{i=1}^{N}(x_i - \mu)^3$	It is a measure of the asymmetry of a distribution around its mean
	Kurtosis	$kurt = \frac{1}{N\sigma^4}\sum_{i=1}^{N}(x_i - \mu)^4$	It is a measure of how different a distribution's tails are from the tails of a normal distribution
	Minimum	$min_{i=1,\dots,n}(x_i)$	Minimum value in a collection of data
	Maximum	$max_{i=1,\dots,n}(x_i)$	Maximum value in a collection of data
	Median	$Med = x_{0.5} : F(x_{0.5}) \leq 0.5$	It is the value separating the higher half from the lower half of a data sample
Frequency	Entropy	$H(x) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$	It estimates the uniformity of signal energy distribution in the frequency domain
	Power	$P(f_k) = \frac{1}{N}\lvert X(f_k)\rvert^2$	It describes how the power of the signal is distributed over frequency
	Highest Frequency	$HF = F(max(x(n)))$	It is the frequency associated to the largest magnitude
Time/Frequency	Number of crossing point	$C(x) = \lvert\{i \in \{1, 2, \dots, n\} : x_i > k\}\rvert$	Given a domain and a threshold k it count all the point over that threshold

50% overlapping windows [17, 61, 72], while, five studies used the overlap rate of 99% [69], 87.5% [34], 75% [59], 25% [91].

With regard to both Time Windowing and Sensor Event Windowing, some studies considered applying an iterative approach by testing multiple window sizes until the optimal one has been defined, to overcome the definition of the window size on an empirical basis. The optimal window size was determined to be 6 steps by Jeong et al. [39] and Nguyen et al. [64], who evaluated window sizes ranging from 1 to 6 steps. Merry et al. [59] evaluated window sizes from 0.5 to 3 seconds, with 1.5 seconds being optimal. D'Arco et al. [17] assessed the window sizes between 1 and 20 seconds, determining the 10 seconds as optimal. Ren et al. [72] experimented with window sizes ranging from 1 to 60 seconds, deciding that 20 seconds was optimal.

3.6 Feature Extraction

Research indicates that with a large number of irrelevant and/or redundant data, it is more possible to run into classifier errors [41]. Accordingly, Feature Extraction is a widely used technique in Machine Learning (ML). It creates a high-level representation of the data segments generated by each sensor, thus allowing the extraction of only the salient data from them and reducing the data dimensionality. From a mathematical point of view, the feature extraction technique can be seen as a set of functions $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ that map each segment $x = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^n$ into a new segment $f_i = g_i(x_1, x_2, ..., x_n)$ for $i = 1, ..., m$ and $m \leq n$.

The choice of g is of vital importance as the result must always respect the nature of the initial data.

Many different features are used in the HAR among the studies included in this review, which are summarised in Table 6. These can be grouped into two main groups, hand-crafted features and learned features. Hand-crafted features were included in 16 studies, in which the features were selected on the heuristic basis of expert opinions, generally using statistical knowledge. Hand-crafted features are generally split into *time-domain* and *frequency-domain* features. The most used hand-crafted features in these articles are summarised in Table 5.

This group of features has been applied for analysing both inertial data and pressure data.

Although they are not shown in the table, it is worth noting other solutions that have been found. In some studies, the pressure sensors have been coupled together to give a measurement of the vertical Ground Reaction Force (vGRF) that has been used either to calculate the difference between the samples according to the body weight [12], or as a mean between samples [36], or raw [13]. Truong et al. [83] used the correlation analysis to identify the relation between axes of the same sensor, such as the accelerometer and gyroscope. Besides statistical features, in [12, 13, 31, 80] the gait analysis has been applied, including detection of foot contact pitch, percentage of double support time and gait phase identification. Furthermore, some singular studies have proposed their features. Jeong et al. [39] extracted the features from each step by applying the following equation on eight pressure sensors:

$$f(i) = \sum_{t=0}^{E_i} p_i(t) * t, \qquad i = [1, 8] \tag{3}$$

where, $p_i(t)$ is the pressure collected from sensor i at sampling time t within step epoch E_i.

Truong et al. [84] introduced three handcrafted pressure features, called Temporally Adaptive Weighting Accumulation (TAWA), that represent the pressure accumulation per step, calculated as the temporal increase, temporal decrease and temporal independent accumulation. The temporal independent accumulation assigns equal weights to each step, whereas, the temporal increase and decrease favour the early and late stages of a step by assigning higher weight.

In learned feature approaches, the features are automatically discovered by means of tools. Multiple approaches can be used, such as Codebooks [76], and to some extent Deep Learning Algorithms. Codebooks consist of the construction of clusters starting from each sensor data window expressed as a sequence. Each cluster centre is called a codeword and represents a statistically distinctive sub-sequence. Then, each sequence is encoded using a bag-of-words approach using codewords as features. However, the research examined did not employ this method. Nevertheless, seven studies preferred to process raw data directly using Deep Learning algorithms, which have the potential to automatically detect patterns within data, while constructing abstract features and classify them [16].

Only one study [47] did not use the feature extraction technique, since it used directly the voltage information generated by the proposed system.

3.7 Feature Selection

Concerning hand-crafted features, all the chosen features are based on heuristic studies, which means that there is no evidence that unnecessary features are avoided. Multiple features can result in irrelevant, misleading or redundant high-dimensional data, increasing the search space size and making data processing more complex [46]. Applying techniques, such as Feature Selection, to evaluate and reduce the features used is a key point for improving the classification performance.

Of the 24 studies, only ten of them used feature selection techniques. They are shown in Table 6.

Sazonov et al. [75] used the Multinomial Logistic Discrimination (MLD) algorithm's forward selection procedure, with the accuracy of the validation dataset as the criterion, for selecting the best features, 12 features, six from the accelerometer and six from the pressure sensors. Differently, De Pinho et al. [5] used Halls's algorithm to select 12 features, four from FSRs and eight from IMU sensors. The Minimum Redundancy Maximum Relevance (MRMR) algorithm was involved in feature selection by Hedge et al. [35], and Chen et al. [12]. The former selected nine features, four from pressure sensors and five from IMU sensors, instead, the latter used the MRMR only to evaluate the features extracted, five, two derived from the pressure sensors and three derived from the IMU sensors. Similarly, D'Arco et al. [17] reduced the number of features from 272 to 227 features by using the univariate selection based on the ANOVA technique. Ren et. al [72] utilised a 20-feature cut-off strategy together with the selection of the number of pressure sensors involved to identify which are the best features in the multiple

Table 6. Summary of the data manipulation activities involved in the included studies

Study	Window Size	Features	Feature Selection Algorithm
Chen et al. [13]	1 stride	Foot contact pith, Foot contact pitch - GRF2 pitch, percentage of double support time	Heuristic (they chose from the available)
	2 s	Total plantar pressure	/
Chen et al. [12]	30 steps	Max GRF difference (in BW), number of threshold crossing points, foot contact pitch, foot contact pitch - GRF2 pitch, percentage of double support time	MRMR (feature evaluation)
D'Arco et al. [17]	From 1 to 20 seconds with and without 50% overlap, best 10 seconds with overlap	Mean, range, standard deviation, skewness, kurtosis, dfr, entropy, energy	Univariate selection using ANOVA
De Pinho et al. [5]	0.3 s	Initial (100): descriptive statistics - standard deviation, variance, minimum, maximum, average value, cumulative difference between samples, Euler angles of pitch, roll and yaw. Selected (12): 2 axes of the gyroscope, 2 axes of the magnetometer, 1 axis of the accelerometer, 4 FSRs, 2 Euler angles and the cumulative difference between samples of the barometer	Hall's algorithm
Dehzangi et al. [22]	/	STFT, Katz, AR, Max Values, Variance, Power, Mean	Evaluation of better sensor, with single sensor classification
González et al. [31]	2.5 s	Mean, largest magnitude in the spectrum, frequency of the largest magnitude, gait phase sequence	PCA
Haescher et al. [34]	17 s with 87.5% overlap	Frequency with the highest amplitude, highest significant frequency, spectral centroid and signal energy	/
Hedge et al. [36]	2 s	Mean of total pressure, standard deviation of total pressure, mean resultant acceleration and standard deviation resultant acceleration	Removed the first 30 seconds of data to steady state
Hedge et al. [37]	2 s	Mean of P_Sum from left/right lower extremity (LE), standard deviation of P_Sum from LE, mean of Resultant Acceleration (RA) from LE, standard deviation RA from LE, number of mean crossing of P_Sum from LE, number of mean crossing of RA from LE	/
Hegde et al [35]	4 s	Mean, standard deviation, number of mean crossing and entropy of sum of pressure and resultant acceleration, average maximum value of superior-inferior acceleration	MRMR
Jeong et al. [39]	From 1 to 6 steps; optimal 6 steps	8 features calculated as: $\sum_{t=0}^{E_i} p_i(t) \cdot t, i = \overline{1, 8}$, where $p_i(t)$ is the pressure collected from sensor i at sampling time t within step epoch E_i	/
Key et al. [40]	15 samples	Standard deviation	
Lan et al. [47]	5 s	Voltage generated by capacitors	/
McCalmont et al. [57]	/	Mean, standard deviation of acceleration, velocity and total acceleration, cadence	/
Merry et al. [59]	0.5, 0.75, 1, 1.5, 2, 2.5 and 3 s, 75% overlap	Mean, mode, median, sum of total of 1's and 0's.	Feature selection with feature ranking: chi-square, Fisher score feature, Gini Index, info-gain, MRMR; Only 10 features are used
Moufawad el Achkar et al. [61]	5 sec with 2.5 overlaps	Raw data	/
Moufawad el Achkar et al. [62]	6 s	Raw data	/
Nguyen et al. [64]	from 1 to 6 steps; optimal 6 steps	skewness-area (SA), pressure area ratio (AR), and kurtosis-area (KA)	/
Paydarfar et al. [69]	1 sec with consecutive windows, means that each sample is the beginning of a new window	Raw data	/

Ren et al. [72]	From 1 to 60 seconds, with 5 seconds off and 50% overlap; optimal one is 20 seconds	General statistics analysis: mean, maximum, standard deviation and median; Peak Analysis: peak number, average and standard deviation of the interval between peaks, average and SD of the peak magnitudes, average and SD of the peak widths; Gait Phase analysis: average of the early stance phase and late stance phase; Frequency Domain Analysis: power density, frequency signal weighted average from 1.67 to 10 Hz, skewness of the frequency components below 10 Hz, mean of the AC components from 2 to 10 Hz, and standard deviation of the same segment; Pressure Distribution Analysis: envelope, anterior-posterior distribution of the plantar pressure, medial forefoot, lateral forefoot	Combination of sensors; feature ranking removing one feature at a time until remaining with only one. The minimum number of features corresponding to the inflexion point for the prediction rate vs. the number of the feature was considered to be the optimum number of inputs. A total of 686 combinations of sensor configurations and number of features were tested (i.e., best 1-sensor: 29, best 2-sensor: 54, best 3-sensor: 76, best 4-sensor: 98, best 5-sensor: 120, best 6-sensor: 142, 7-sensor: 167)
Sazonov et al. [75]	2 s	Mean, entropy, standard deviation	Forward selection procedure applied to the MLD classification model with the accuracy of the validation dataset as criterion
Song et al. [80]	/	Average clustering coefficient, characteristic path length, clustering coefficient entropy, and path distribution entropy	/
Truong et al. [83]	10 s	Features from inertial data: mean, median, mode, range, skewness, kurtosis, 4-th and 5-th central moment, standard deviation, variance, mean absolute deviation, and sum of absolute values, mean absolute deviation (MAD), signal magnitude area (SMA), root mean square (RMS), intensity of movement (IM), sum of absolute values (SAV), correlation between axes values, average energy (AE), dominant frequency (DF), and amplitude values. Features from pressure data: mean, max. standard deviation of pressure data; Correlation between the counterpart sensors from both feet; Pressure Area	Feature score using RELIEF-F, k=10. Total number of features selected 20
Truong et al. [84]	From 1 to 8 steps, focusing only to 1 and 2 steps at the end	Temporally adaptive weighting accumulation, standard deviation	/
Wang et al. [85]	6 s	Raw data	/
Zhang et al. [91]	2 s, 25% overlap	Raw data	/

AR: auto regressive parameters; BW: body weight; GRF: ground reaction force; MRMR: minimum redundancy maximum relevance; PCA: principal component analysis; STFT: short time fourier transform.

experiments. This method helped to minimise the number of features from 140 to 44. Merry et al. [59] selected ten features, analysing the results of five distinct feature ranking techniques, including Chi-square, Fischer score feature, Gini index, Info-gain, and MRMR. Dehzangi et al. [22], instead of focusing on the smallest amount of features, employed a ranking technique to choose the number of pressure cells to use. The ideal number of pressure cells was determined to be nine. A feature scoring approach named RELIEF-F was used to select the ideal set of features by Truong et al. [83], highlighting an optimal subset of 20 features. Gonzalez et al. [31] used the Principal Component Analysis (PCA) to reduce the feature number. The PCA identifies the eigenvectors of the covariance matrix with the highest eigenvalues and then uses those to project the data into a new subspace of equal or fewer dimensions. In this study, the number of features was reduced from 37 to 22.

3.8 HAR Algorithms and Validation

Human Activity Recognition algorithms can be broadly categorised into model-driven and data-driven approaches. In the model-driven approaches, explicit representations and rules are defined through a heuristic analysis by an expert. In the data-driven approaches, activity recognition is made on the basis of data already in possession, which is processed by an algorithm that generates a sequence of conditional statements.

This section will summarise the HAR algorithms used in the 24 selected studies. Table 7 reports the information related to the solution analysed, including algorithms, dataset balancing, validation techniques and evaluation metrics.

Of the 24 selected studies, only two studies [61, 62], both proposed by Moufawad et Achkar et al., used a model-driven approach for activity recognition. They proposed an event-driven classification tree for the classification

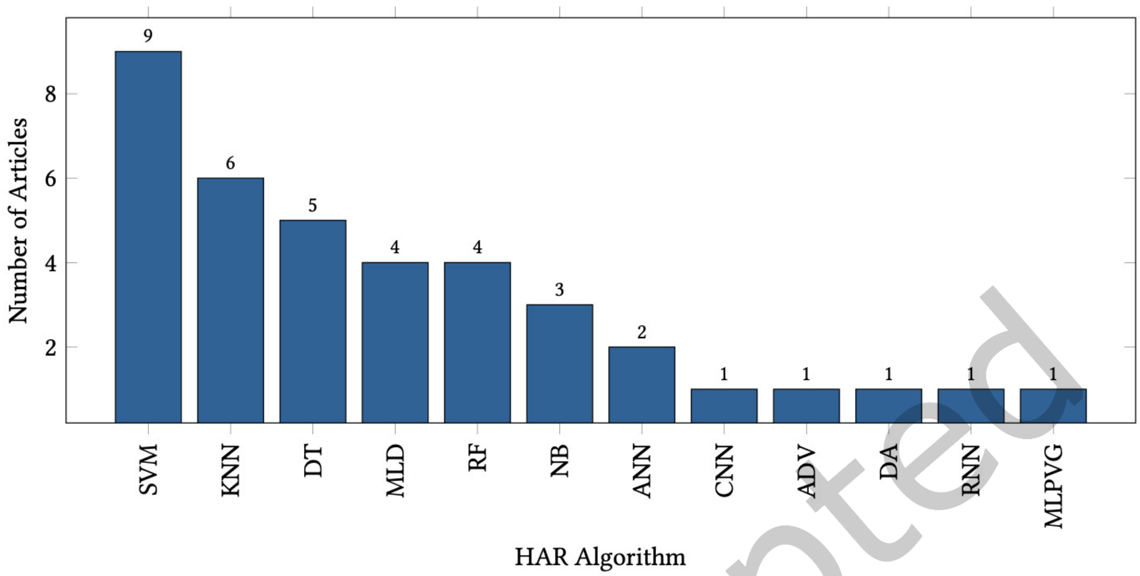


Fig. 6. Use of machine learning algorithms in the included studies.

of the activities using data from instrumented shoes. To distinguish between locomotion and non-locomotion activities, pitch angular velocity was used. The total force of the pressure sensors distinguished sitting from standing in non-locomotion activities; however, in standing, a distinction was made using barometer values between simple standing, elevator up, and elevator down. The barometer readings were utilised to identify level walking against downstairs walking and upstairs walking in locomotion activities, which were then distinguished with the equivalent downhill and uphill activities using the foot angle, calculated from the accelerometer.

The rest, 23 studies, relied on a data-driven approach, and consequently on ML algorithms. The ML algorithms used in the selected articles are reported in Fig. 6.

The Support Vector Machine (SVM) was the most widely used ML algorithm, with nine studies using it. SVM is a kernel-based algorithm that creates hyperplanes in order to determine the class to which each sample belongs. Among the analysed studies several kernels have been applied, including linear [13, 17] and radial [59, 75]. The K-Nearest Neighbours (KNN), following the SVM, was the most used, appearing in six studies. This algorithm makes its strength the ability to be a lightweight algorithm to use and is very versatile. It determines the class to which a sample belongs by calculating its distance to its nearest neighbours. The number of neighbours (k) can be specified, so that in the analysed studies there was the use of k ranging from 1 to 9, usually using odds numbers, except for the study proposed by Nguyen et al. [64] in which the k was equal to 2. The Decision Tree (DT) was employed in five studies [39, 40, 47, 59, 64]. It creates classification rules by connecting the internal nodes, in which the features are tested, and leaf nodes, which represent the class labels. An improvement of the DT is the Random Forest (RF), which is an ensemble algorithm that uses multiple DTs to determine the belonging class to a specific sample. Four studies used it [5, 47, 57, 72], in which the number of DTs was limited to 100. Four studies [35–37, 75] opted for the use of the Multinomial Logistic Discrimination (MLD), which combines a logistic regression model that estimates the probability of a class member with a decision rule that renders the projected probabilities of the outcome into a categorical output. Naive Bayes (NB) algorithms, which are a family of algorithms based on Bayes theory, were similarly used in three studies. González et al. [31] used a Gaussian

Naive Bayes algorithm, while Hascher et al. [34] preferred the use of a Bayesian Network. Lan et al. [47] did not specify the type of the NB algorithm. The Discriminant Analysis (DA), which produces a linear combination of features that describes or differentiates two or more classes, has only been used in the study proposed by Merry et al. [59]. Chen et al. [13] presented a hybrid solution that integrates both ML and threshold algorithms. They distinguish between dynamic activities like walking and running and quasi-dynamic activities like sitting and standing. The first category is classified by an SVM, while the second category is managed by a threshold algorithm that considers the amount of pressure applied to the insole.

Deep learning is becoming a viable alternative to ML systems as a result of recent technical advancements, as it can autonomously discover salient features within raw data and utilise them to classify activities, avoiding feature extraction and selection [33]. Deep learning algorithms try to replicate how the neurons in a human brain process information. Deep learning algorithms are made up of layers in which data are processed and filtered before being used to make a prediction. Seven of the studies analysed opted to use deep learning for their solutions. McCalmont et al. [57] used an Artificial Neural Network (ANN), while, Sazonov et al. [75] developed a Multi-Layer Perceptron (MLP) composed of one hidden layer with hidden four neurons, with sigmoid and linear activation functions on the hidden and output layers, respectively.

The nodes in MLPs are usually connected to all neurons in the next layer, resulting in fully connected networks that are prone to overfitting. An evolution of such an algorithm is the Convolutional Neural Networks (CNNs), which take advantage of the hierarchical structure in data and create patterns of increasing complexity utilising smaller and simpler patterns embossed in their filters. Wang et al. [85] proposed a CNN consisting of one-dimensional layers in which the information is handled by squeezing the data in the sample. Paydafar et al. [69] presented a CNN embedding a Recurrent Neural Network (RNN). The RNN allows introducing backward connections so that each pattern can be assumed to be dependent on previous ones. Zhang et al. [91] proposed an adversarial neural network consisting of a feature extractor, an activity recogniser and a subject discriminator. The feature extractor is created using a CNN that extracts the salient information from the data, which, in turn, is used by a fully connected network to determine the activity performed. The information retrieved from CNN and the label received from the activity recogniser are both inserted into a third neural network that determines which subject generated the information. This architecture was provided to detect data that is unrelated to study participants by offering a system that can operate independently of the people included in the research data collection.

Song et al. [80] presented a network construction method based on multi-layer LPVG (MLPVG). They developed a two-layer architecture. The first layer converts 21 pressure sensors' data into a Limited Penetrable Visibility Graph (LPVG) that obtains the characteristic path length of the different gait periods. The second layer, using the characteristic path length, produces the joint distribution of the average clustering coefficient and the maximum degree. The network obtained is used to differentiate between activities.

Data-driven algorithms are usually engineered with the assumption that data distribution is balanced across classes. If this assumption is violated, the performance of the algorithm may be skewed towards classes having majority samples [54]. In order to improve the robustness of the solutions, the problem of imbalanced datasets should be identified and solved. Among the studies analysed, only three considered the issue of balancing the dataset. Chen et al. [12, 13] applied a downsampling technique, reducing the number of samples to the same number for each class. Ren et al. [72] reported that some participants abandoned their experiment, thus they decided to use just the data acquired by those who finished the entire set of activities for training and the rest for testing, resulting in a balanced dataset for training. There is no mention of managing the balancing of the dataset in the other works; nonetheless, an analysis of how the datasets were constructed revealed that 12 studies employed a balanced dataset and six studies did not [5, 17, 35, 47, 75, 85]. Furthermore, because they used a model-driven algorithm, Moufawad el Ackhar et al. [61, 62] used their datasets only for testing purposes. There

Table 7. Summary of the algorithms with respective performance involved in the included studies

Study	Algorithm	Balanced dataset	Validation	Evaluation Metrics (%)				
				Accuracy	Precision	Sensitivity	Specificity	F1-Score
Chen et al. [13]	SVM	Yes	5F-CV	99.8	/	/	/	/
	Threshold-based	Yes	/	/	/	/	/	/
Chen et al. [12]	SVM	Yes	5F-CV	98.1	98.12	98.11	99.62	98.11
D'Arco et al. [17]	SVM	No	5F-CV	94.66	95.09	94.66	/	94.64
De Pinho et al. [5]	RF	No	LOSO-CV	93.34	/	/	/	/
Dehzangi et al. [22]	SVM	Yes	10F-CV	97.6	/	/	/	/
González et al. [31]	GNB	Yes	6F-CV	92	/	/	/	/
Haescher et al. [34]	BN	Yes	LOSO-CV	/	/	86.17	/	/
			10F-CV	74.85	74.48	74.85	/	74.55
Hedge et al. [36]	MLD	Yes	LOO-CV	96.9	97.25	97	/	/
Hedge et al. [37]	MLD	Yes	LOSO-CV	Healthy: 96.2; CP: 95.3	Healthy: 96; CP: 95.67	Healthy: 96; CP: 94.67	/	/
Hegde et al [35]	MLD	No	LOSO-CV	10 classes: 90; 13 classes: 81.15	13 classes: 82.37	13 classes: 81.15	/	13 classes: 81.56
Jeong et al. [39]	DT	No	LOO-CV	87.6	/	/	/	/
	KNN			91.4	/	/	/	/
	SVM			95.2	95.2	95.2	97.6	95.23
Key et al. [40]	DT	Yes	/	76	/	/	/	/
Lan et al. [47]	NB	No	10F-CV	95.08	/	/	/	/
	KNN			93.83	/	/	/	/
	DT			94.57	/	/	/	/
	RF			94.87	/	/	/	/
McCalmont et al. [57]	ANN	/	80% training 20% testing	80	/	80	80	
	RF			70	/	70	73.33	/
	KNN			70	/	80	63.33	/
Merry et al. [59]	SVM	Yes	LOSO-CV	99.03 ± 0.82	/	/	/	/
	DT			99.07 ± 0.76	/	/	/	/
	KNN			99.07 ± 0.86	/	/	/	/
	DA			99.11 ± 0.56	/	/	/	/
Moufawad el Achkar et al. [61]	Threshold-based	No	/	99.03 ± 0.82	95.03	93.01	99.60	93.84
Moufawad el Achkar et al. [62]	Threshold-based	No	/	93	92.73	96.48	92.33	92.48
Nguyen et al. [64]	KNN	/	LOO-CV	97.84	97.99	97.84	/	97.83
	SVM			87.66	/	/	/	/
	DT			95.40	/	/	/	/

Paydarfar et al. [69]	RNN	Yes	LOSO-CV	87.0 ± 8.9	/	/	/	/
Ren et al. [72]	RF	Yes	6 subjects for the training (the ones that completed all the activities), 11 for the testing. 33% train split repeated 20 times	89	/	/	/	/
Sazonov et al. [75]	SVM	No	LOSO-CV	97.9	94.9	95.5	98.4	95.2
	MLD			97.5	92.3	94.7	98.1	93.4
	MLP			97.9	93.7	95.0	98.3	94.3
Song et al. [80]	MLPVG	Yes	/	93.91	/	/	/	/
Truong et al. [83]	KNN	Yes	/	80	81.8	82.5	/	81.7
Truong et al. [84]	SVM	Yes	5F-CV	1 step: 83.55; 2 steps: 89.87; 8 steps: 97.29	1 step: 83.04; 2 steps: 89.58	1 step: 83.34; 2 steps: 89.89	/	1 step: 83.18; 2 steps: 89.72
Wang et al. [85]	1D-CNN	No	/	98.61	/	97.92	99.58	/
Zhang et al. [91]	Adversarial Network	Yes	LOSO-CV	98.92	/	/	/	/

1D-CNN: one-dimensional convolutional neural network; ANN: artificial neural network; BN: bayes net; CapsNet: capsule network; CNN: convolutional neural network; CP: cerebral palsy participants; DA: discriminant analysis; DT: decision tree; xF-CV: x-fold cross validation; GNB: gaussian naive bayes; KNN: k-nearest neighbour; LOO-CV: leave-one-out cross validation; LOSO-CV: leave-one-subject-out cross validation; MLD: multinomial logistic discrimination; MLP: multi layer perceptron; MLPVG: multi-layer limited penetrable visibility graph; NB: naive bayes; RF: random forest; RNN: recursive neural network; SVM: support vector machine

is insufficient information about the dataset used in three research [39, 57, 64], making it impossible to establish if they employed a balanced dataset or not.

Different strategies have been utilised to train and evaluate the algorithms in order to produce more accurate and reliable results. Eight studies used k-fold cross-validation (kF-CV), which splits the dataset into equal chunks and uses one for testing and the other for model training. Three studies using a k of 10 [22, 34, 47], four articles using a k of 5 [12, 13, 17, 84], and one article using a k of 6 [31]. In 3 studies [36, 39, 64], the leave-one-out cross-validation (LOO-CV) was used, which used all the samples but one for the training and one for testing. Eight studies used the leave-one-subject-out cross-validation (LOSO-CV), which reserves all the data of a participant for the testing while the others are used for training. Another validation technique that can be used is a priori division of the dataset into several sections in order to use each section in a certain phase, for example, training and testing. McCalmont et al. [57] and Ren et al. [72] did a split between the training and test set, using a cut for the training set of 20% and 33%, respectively. In addition, Ren et al. [72] applied a data selection based on participants, including six participants' data for training and the rest for the testing, repeating the training 20 times. Furthermore, Moufawad el Achkar et al. [61, 62] used the entire dataset for testing purposes, since a threshold-based algorithm was employed. Four studies [40, 80, 83, 85] did not specify the validation technique used in their studies.

3.9 HAR Performance Assessment

To evaluate the performance of the HAR classifiers used in the different studies, five heuristic metrics (accuracy, precision, sensitivity, specificity and F1-score) were chosen, based on those used by the studies analysed. These metrics are extracted from the confusion matrix, which is an error matrix that compares the ground truth (the observed labels) with the estimated labels [81]. From the confusion matrix, four main values can be extracted and later used to compute the metrics: *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, and *False Negative (FN)*. The *TP* is the number of predictions where the classifier correctly predicts the positive class as positive, the *TN* is the number of predictions where the classifier correctly predicts the negative class as negative, the *FP* is the number of predictions where the classifier incorrectly predicts the negative class as positive, the *FN* is the number of predictions where the classifier incorrectly predicts the positive class as negative.

Accuracy is the percentage of correctly predicted data points out of all the data points, in other words, it measures how often the algorithm classifies a data point correctly (see Eq. 4).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{4}$$

Unfortunately, *Accuracy* has downsides such as the possibility of being compromised by the balancing of the dataset used. So it needs to be flanked by other metrics in order to have a more robust validation.

Precision is the number of correct positive results divided by the number of positive results predicted by the classifier (see Eq. 5).

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Sensitivity is the proportion of the number of correct positive results out of the number of all positive samples (see Eq. 6).

$$Sensitivity = \frac{TP}{TP + FN} \tag{6}$$

Specificity is the proportion of the number of correct negative results out of the number of all negative samples (see Eq. 7)

$$Specificity = \frac{TN}{TN + FP} \tag{7}$$

F1_Score is the Harmonic Mean between precision and recall. It tells you how precise your classifier is, as well as how robust it is (see Eq. 8).

$$F1\_Score = 2 \times \frac{Precision * Recall}{Precision + Recall} \tag{8}$$

The recognition of human activities is mainly aimed at the classification of multiple activities, therefore, it is essential to consider this aspect when evaluating the results obtained from a solution. In multiclass classification, mainly in situations where the dataset is imbalanced, using the above metrics does not always bring the actual results, as the number of samples for each class is not considered. For this reason, we speak of class metrics, i.e. the multiclass classification problem is transformed into a problem where each class is compared with the rest and the metrics are calculated. To obtain a result that can be meaningful for the whole solution, the metrics of each class can be merged mainly using two strategies, macro or weighted metrics. The first combines the metrics of each class using the arithmetic mean, while the other combines them taking into account the number of samples per class, therefore a weighted average.

When assessing the outcomes of a solution, it is crucial to keep in mind that the recognition of human actions is primarily focused on the classification of numerous activities. A classifier is built generally with the assumption that the samples are equally distributed between the classes. If this assumption is violated the effectiveness of the evaluation metrics can be altered. For this reason, transforming a multiclass classification problem into a problem where each class is compared against the others and computing the metrics for each class can be a better approach. The metrics of the classes can be combined using two strategies, macro or weighted metrics, to produce a result that can be relevant to the entire solution [27]. The first combines the metrics of each class using the arithmetic mean, while the other combines them taking into account the proportion of the number of samples per class relative to the total number of samples. Weighted average metrics shall be considered when a dataset is imbalanced.

Accuracy is the most popular measurement used to assess performance. Twenty-five out of the twenty-six studies used it. Overall *Accuracy* of those studies varies from 70% [57] to 99.8% [13]. Fourteen out of twenty-six studies achieved an Accuracy greater than 95%.

The other metrics are not always presented in the studies, but if a confusion matrix was provided in the article, they have been extracted for comparison with other solutions.

The *Precision* has been extracted from 13 articles, the lowest achieved is 74.48% [34] whereas the highest is 98.12% [12]. Seven articles presented the *Specificity* value, with a range between 63.33% [57] and 99.62% [12], whereas, 15 articles provided the *Sensitivity* value ranging from 74.85 [34] to 98.11% [12]. The *F1-Score* was highlighted in 12 articles, the lowest value is 81.7% [83] and the highest value is 98.11% [12].

3.10 Quality Assessment

The selected articles were evaluated based on the quality assessment method proposed by Kmet et al. [44]. The assessment was carried out by Reviewer 1 and Reviewer 2. The Total Score and the Quality Score were calculated using Eq. 1.

The quality assessment results highlighted that nine articles were High Quality (HQ), where three studies of them satisfied completely all the items obtaining a score of 100% [37, 47, 62]. Among the remaining, 12 of them fell into Good Quality (GQ) and five into Moderate Quality (MQ). No article was identified as Low Quality (LQ). The results of the quality assessment are shown in Table 8.

4 DISCUSSION

In this systematic review, 419 studies were screened using a five-database search strategy, however, only 26 research supplied adequate information to be included, in order to establish the potential of employing smart insoles for the recognition of Activities of Daily Living (ADLs). Although multiple reviewers participated in the analysis of the articles, to avoid possible bias in the judgements of the latter, a quality assessment was carried out on the selected research, comprising the entire structure and definition of the solutions, revealing that no article was classified as low quality.

Analysing the findings of the review, the current challenges and gaps, that should be addressed in future studies, have been identified and summarised as follows:

(1) According to the inclusion criteria, all types of ADLs were included in this review. However, upon analysis of the identified articles, it was noted that ambulation and fitness-related activities were the most frequently explored activities in the research. This is unsurprising, as these activities heavily involve the lower limbs, which is where smart insoles provide the most benefit. Although four studies examined complex activities, including vacuuming [35], tooth brushing [83], slip hazard [12] and falls [85], the smart insoles achieved less satisfactory performance in these cases. Overall, the performance achieved in the included studies ranged from 70% to 99.8% in *Accuracy*, with 15 studies having an *Accuracy* greater than 95%. Hedge et al. [35] compared the performance of smart insoles to that of a wristband for both ambulation and complex activity recognition. Their findings indicated that smart insoles outperformed wristbands by approximately 10%. Similarly, Duong et al. [23] reported that smart insoles were reliable for identifying six common ambulation activities when used as a stand-alone device, surpassing the performance of wristbands. In contrast, Truong et al. [83] determined that wristbands outperformed smart insoles. However, it is important to note that Truong et al. primarily analysed complex activities in which the upper body played a more significant role. All three studies [23, 35, 83] found that the optimal performance for recognising complex activities was achieved when the smart insoles were used in conjunction with an additional device located in the

Table 8. Quality assessment results for the included studies

Study	**Items**														**Scores**		**Category**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Tot**	**%**	
Chen et al. [13]	2	2	N/A	2	2	1	1	2	1	2	1	2	1	2	21	75	MQ
Chen et al. [12]	2	2	1	2	2	N/A	1	2	2	2	2	2	2	2	24	85.71	GQ
D'Arco et al. [17]	2	2	1	2	2	2	1	2	1	2	2	2	2	2	25	89.29	GQ
De Pinho et al. [5]	2	2	1	2	2	0	2	2	2	2	1	2	2	2	24	85.71	GQ
Dehzangi et al. [22]	2	2	1	2	2	1	2	2	2	2	2	2	1	2	25	89.29	GQ
González et al. [31]	2	2	2	2	2	2	1	2	2	2	2	2	2	2	27	96.43	HQ
Haescher et al. [34]	2	2	2	2	2	1	2	2	2	2	2	2	2	2	27	96.43	HQ
Hedge et al. [36]	2	2	2	2	2	2	N/A	2	2	2	1	2	2	2	25	89.29	GQ
Hedge et al. [35]	2	2	2	2	2	2	N/A	2	2	2	2	2	2	2	26	92.86	GQ
Hegde et al [37]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	28	100	HQ
Jeong et al. [39]	2	2	1	2	N/A	1	2	2	1	2	2	2	2	2	23	82.14	MQ
Key et al. [40]	2	2	N/A	2	2	1	1	2	1	2	1	2	2	2	22	78.57	MQ
Lan et al. [47]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	28	100	HQ
McCalmont et al. [57]	2	2	N/A	2	2	2	N/A	2	2	2	1	2	2	2	23	82.14	MQ
Merry et al. [59]	2	2	1	2	2	2	1	2	2	2	2	2	1	2	25	89.29	GQ
Moufawad el Achkar et al. [61]	2	2	1	2	2	2	2	2	1	2	0	2	2	2	24	85.71	GQ
Moufawad el Achkar et al. [62]	2	2	2	2	2	2	2	2	2	2	2	2	2	2	28	100	HQ
Nguyen et al. [64]	2	2	2	2	2	2	2	2	1	2	2	2	2	2	27	96.43	HQ
Paydarfar et al. [69]	2	2	1	2	2	2	N/A	2	2	2	2	2	1	2	24	85.71	GQ
Ren et al. [72]	2	2	2	2	2	2	2	2	2	2	2	2	1	2	27	96.43	HQ
Sazonov et al. [75]	2	2	2	2	2	2	N/A	2	2	2	1	2	2	2	25	89.29	GQ
Song et al. [80]	2	2	2	2	1	2	2	1	2	2	2	2	2	2	26	92.86	GQ
Truong et al. [83]	2	2	N/A	2	2	1	2	2	N/A	2	2	2	2	2	23	82.14	MQ
Truong et al. [84]	2	2	1	2	2	2	2	2	2	2	2	2	2	2	27	96.43	HQ
Wang et al. [85]	2	2	2	2	2	2	N/A	2	2	2	1	2	2	2	25	89.29	GQ
Zhang et al. [91]	2	2	2	2	2	2	2	2	2	2	2	2	1	2	27	96.43	HQ

upper body. This suggests that solutions that incorporate an upper body device are preferred to accurately recognise complex activities. In addition, Duong et al. [23] analysed multiple subgroups using wristbands, smart insoles, and smartphones. Their analysis revealed that there were no statistical differences between the subgroups, confirming the previous assertion that smart insoles require an additional device in the upper body for recognising complex activities regardless of the type.

(2) This systematic review focused only on studies that used a smart insole as the only data collection device. Different kinds of sensors, ranging from pressure sensors to IMU sensors, were identified in the studies examined. Nine of the studies encapsulated all the required sensors inside the smart insole, whereas, the others adopted an auxiliary system attached to the shoe (in the back, top or side) or ankle, where inertial

sensors, ECU and battery were inserted. Although integrating everything inside the insole considerably reduces the footprint for the user and allows daily use without any hindrance, its duty cycle is considerably lower than the counterpart with an auxiliary system mainly due to the smaller battery capacity resulting from space limitations. However, even the auxiliary system, which does not involve impediments in walking for the user [7], has disadvantages in terms of usability and acceptability by the user as it is more invasive. Thus, the studies reviewed suggested that there is a trade-off between performance and acceptability that should be considered. Among the studies analysed, only Hedge et al. [35] conducted a study aimed at assessing the acceptability and usability of smart insoles, compared to an inertia sensor and a wrist sensor. The study involved 15 participants who used these devices. The results indicate that the degree of acceptability for the smart insoles is significantly higher compared to the other devices. Additionally, the level of anxiety associated with using the smart insoles was significantly lower than the other devices. The findings suggest that participants viewed the smart insoles as regular insoles, and they did not feel any pressure from being monitored. However, one limitation of the smart insoles is that they can only be used when the user wears shoes. Therefore, they may not be used for the entire day. Overall, the study provided evidence that smart insoles are highly acceptable and usable for individuals who need to monitor their physical activity.

(3) Using smart insole as the only device has provided high performance in all the studies analysis, nevertheless, it has not been possible to identify which sensor is most reliable for recognising a specific activity. Among those analysed, a few articles carried out an evaluation of the positioning of the sensors (referring to pressure sensors) and the suitable number to be inserted into such systems. According to Ren et al. [72], when considering only a single pressure sensor the best location is the heel, whereas in a three-sensor configuration the heel, lateral midfoot, and centre of the forefoot are the optimal locations. D'Arco et al. [17] analysed the importance of both pressure sensors and inertial sensors by assessing the number of features selected for each sensor. The results highlighted that for pressure sensors the most important locations were the hallux, the arch and the heel, whereas, for the inertial sensors, the accelerometer stands out among the inertial sensors, particularly on medio-lateral and anterior-posterior axes. Dehzangi et al. [22] determined that the most important location for pressure sensors is the upper part of the heel. Although these studies have provided fascinating results, highlighting the importance of the heel location for pressure sensors, an in-depth analysis of the minimum configuration of sensors required to recognise specific activities is missing and requires further investigation.

(4) Collecting, validating and labelling data is a time-consuming task. Among the studies analysed, the solution used ranged from manual inspection and validation to automating validation and labelling via smartphone apps, with some cases where data was validated via video. Manually inspecting and validating data allows for a high level of precision and accuracy in the labelling process, but results in time-consuming and expensive, especially when dealing with large datasets. Additionally, manual labelling may be subject to bias and human error, which can affect the quality of the data [88]. Automated approaches using smartphone apps, instead, can be more successful and economical and might be used in uncontrolled environments; nevertheless, the accuracy of the labelling could be impacted by the software's constraints and the setting in which the data is collected. Videos-based validation approaches can provide more insights and context about the data and can help identify anomalies and inconsistencies. However, it can be time-consuming and resource-intensive and requires the data to be collected in a controlled environment. Using multi-device sessions, including silver/gold standard device [23], could provide insightful information and could favour the data collection without any supervision, however, it could result in discomfort for the user, who can be reluctant to use the system in daily living. While there is no universal solution to the problem, it has been suggested that a combination of multiple techniques can be applied to optimise the validation and labelling process especially when the data is collected in uncontrolled environments.

(5) The smart insoles, given their small size, are not equipped with a processing unit capable of processing the collected data, for this reason, the latter is shared using transmission protocols to more performing devices, such as computers or smartphones. The most used protocol is Bluetooth. However in indoor usage, all wireless protocols are valid, and they differ from each other only in terms of coverage and energy expenditure. In the studies analysed, only one study [34] preferred to use a wired connection between smart insoles and the computer which is great for research purposes but makes it completely unusable in an everyday context.
(6) In the studies analysed, the sampling frequency in the data collection was defined on a heuristic basis, and each researcher determined the most favourable for their objectives. Although the most used sampling frequency was 50 Hz, there was no study that justified the advantages of this choice. In a study published by Liu et al. [52], it was claimed that a reduced sampling rate improved performance, however, the validity of this study for smart insoles has not been verified yet.
(7) Activity classification is influenced by data segmentation. Generally, heuristic decisions were made in the evaluated research to specify the size of the window with which to segment the data. However, because this approach does not guarantee an optimal outcome, in-depth studies on window size, such as those conducted by D'Arco et al. [17], Ren et al. [72], Merry et al. [59] for Time Windowing, and Nguyen et al. [64] and Jeong et al. [39] for Sensor Event Windowing, should be conducted to enhance classification. Furthermore, overlapping of the windows should be considered since it can discover boundary case activity in time series streams.
(8) Data-driven solutions are the most common solutions found in the studies examined. To find patterns that differentiate the various activities, this type of solution necessitates a large number of samples. The robustness of the constructed dataset is affected by the number of participants included in the collection of activities. The average number of participants in the studies examined was ten, which appears to provide adequate differentiation. In the literature, however, there are few publicly available HAR datasets that involve smart insoles. Only De Pinho et al. [5] disclosed the dataset used, among the studies examined. Hence, in future, it may be beneficial to create a public dataset that can be used to compare multiple solutions, as now each study focuses on its own dataset with different sensors and placement.
(9) The solutions analysed mainly employed machine learning models to process data and determine the activities conducted by an individual. The most used algorithm was the SVM which also achieved the best performances with an accuracy of 99.8% [13]. In particular, among the performances observed, generally, the shallow machine learning algorithms performed better or equal to those of deep learning. However, this consideration needs further analysis, since only two studies have compared the two types of solutions, in particular, McCalmont et al. [57] reported that an ANN performed better than an RF and a KNN, while, Sazonov et al. [75] reported that an SVM outperformed an MLP. Nevertheless, both solutions have drawbacks. In shallow machine learning, features need to be extracted, but a statistical and mathematical analysis by an expert is required, which takes a long time. In deep learning, on the other hand, raw data can be directly processed, but a large amount of data is required to obtain reliable results. Furthermore, deep learning can be seen as a black box in which an input is provided and output is retrieved without knowing the contribution of each sensor/characteristic unless highly advanced techniques are used.
(10) Machine learning algorithms are usually created with the assumption that there are a number of balanced samples for each class, meaning that if the distribution of the classes in the dataset is skewed or biased, it can be a challenge for predictive modelling. The imbalanced dataset issue appears to have been underestimated in the studies analysed as only three papers have taken this problem into account [12, 13, 72], while 12 other articles have implicitly used a balanced dataset as a result of their data collection phase. The use of re-weighted loss functions to emphasise the importance of minority class samples during training is an alternative solution to oversampling and undersampling [24]. Without altering the initial dataset, such

a technique can help to address the issue of class imbalance and improve the performance of predictive models. This approach, however, was not used in any of the studies examined.

(11) Estimating a HAR system's ability to identify specific activities is critical for determining the validity of a proposed solution. Multiple algorithm validation approaches have been used in the analysed studies. Cross-validation approaches are the most used. It is reasonable to consider that k-fold cross-validation is better suited to large datasets and that it allows for the creation of multiple datasets both for testing and for training, however, if the data are extracted from the same people there may be equality between the subsets and therefore falsify the results obtained. Furthermore, the leave-one-out cross-validation is a subset of the k-fold preferably used in small datasets, since it leaves only one sample for testing. The optimal strategy should employ a leave-one-subject-out, which enables the algorithm to be tested on data from subjects who were not included in the training, demonstrating the robustness of the suggested solution. Additionally, it is wise to consider maximising the number of participants as the evaluation of the algorithm may be affected by the same participant repeatedly performing the same activities for a long time because the data used to test and train the algorithm will come from the same subject, which would allow to look into corner cases and develop benchmark problems for real-life applications. Assessing a solution's performance and dependability by using evaluation metrics is a viable choice. Metrics taken from the confusion matrix are the most commonly utilised solution. Despite the most used metric being *Accuracy*, it is advisable not to use that alone, as it can be affected by the structure and balance of the dataset, but to use other metrics such as *Precision*, *Specificity*, *Sensitivity* and *F1-Score*. Furthermore, the extraction of metrics per class in multiclass classification is recommended, as they allow a better understanding of the behaviour of the solution with respect to each class, and from these obtain the overall metrics using a macro or weighted strategy according to the needs of the experiment. Although these are the most used metrics, it is worth noting that other metrics such as the Area Under the Curve (AUC) might be employed, which represents the degree or measure of separability of the ROC curve and indicates how well the model is able to distinguish between classes. However, it has never been used in the reviewed studies.

(12) Smart insoles have multiple advantages in terms of usability for the user. However, it's important to consider that their performance can be affected by multiple factors when used for extended periods of time. In controlled environments, which have been identified as the most commonly used in analysed studies, these factors can be minimised and kept under control. However, when they are used in free-living scenarios, in other words, uncontrolled environments, these factors must be taken into account. The battery life can limit their use, and data collection methods, such as using an external device like a smartphone, require both data transmission and storage technologies to be refined, not to mention the constraints they pose for the user. Additionally, the conditions under which smart insoles are used can alter the data collected, such as the temperature and humidity inside the shoe, the footwear of the person, and the shoe's structure. Similarly, walking surfaces can also result in altered data, such as when walking on grass, carpet, uneven surface, and inclined surface [56]. Weather and environmental conditions like rain and snow can further affect the proper functioning of smart insoles. Analysing and evaluating the impact of these factors on activity recognition as well as on user comfort should be one of the main focuses of future studies for better integrating such systems into everyday life applications.

(13) Smart insoles can be integrated into several healthcare systems, according to this systematic literature review. The development of a system that was unobtrusive for the user was the common goal of all the studies analysed; nevertheless, a wide range of applications have been reported ranging from telemedicine to rehabilitation for monitoring daily activities. This kind of assessment enables the development of insights over time, highlighting, for instance, variations in mobility [62] and enabling the identification of gait patterns for each activity [13]. The daily energy expenditure and quality of life of an individual can also be assessed through ADLs monitoring [35]. Differentiating between routine everyday activities and

occurrences of falls [85] or slips [12] can be essential to improve the accuracy of the predictions and to apply recovery procedures in at-risk populations, such as the elderly or in industrial situations. Smart insoles can be employed in the medical domain to evaluate patients' progress or rehabilitation by keeping track of the ADLs they accomplish. Merry et al. [59] offered a system for determining the connection between weight-bearing fasciitis and plantar fasciitis, while Hegde et al. [37] developed a system for monitoring children with cerebral palsy.

Overall, the smart insoles have shown promising results from this systematic review. They achieved high performance while minimising the encumbrance for the user. Smart insoles enable continuous physiological parameter monitoring in a non-invasive and autonomous manner throughout a person's daily life. It considerably minimises, if not eliminates, user privacy problems because, unlike other systems, such as image-based solutions that acquire videos, they only collect sensor data and can be inserted directly inside a shoe. Smart insoles don't need any installation and are not tied to a predefined environment, making their usage easier for an individual, as well as reducing the overall cost of the system.

5 CONCLUSION

This systematic review analysed the existing articles in the literature and evaluated the capability of smart insoles as the only device for the recognition of Activities of Daily Living (ADLs).

The review strategy followed the PRISMA guidelines, taking into account only studies in which the smart insoles and ADLs have been included. An in-depth examination of the many aspects of the development of a smart insole-based HAR system has been completed, including the sensing elements used, the participants involved in the studies, the activities recognised, the data segmentation, the feature extraction, the feature selection, the HAR algorithms, the validation strategy, and the performance assessment metrics.

The results obtained from the review revealed the capabilities of smart insoles in recognising human activities, as long as they involve the lower part of the body, such as ambulation and fitness activities. For activities in which only the upper part of the body is involved, the use of only smart insoles is not sufficient to achieve adequate performance, and additional devices, such as wrist sensors are required.

Although the performance achieved in almost all studies is promising, as most of the selected articles have achieved excellent performance, ranging from 70% to 99.8% of *Accuracy*, with 13 studies over 95%, some limitations have been highlighted. There is a lack of studies on the importance of individual sensors and data segmentation techniques, as well as an absence of publicly available datasets including smart insoles that can be used as benchmarks. The problem of the imbalanced dataset has not been adequately addressed by the selected studies, which included mostly data-driven algorithms, and the *Accuracy* as the main evaluation metric, which can be skewed towards the majority class of the dataset, compromising the results. To improve the reliability of the proposed solutions future research should take into account per-class metrics that better deal with imbalanced datasets and evaluate the solution performance by using data not previously seen by the classifier such as using leave-one-subject-out cross-validation. Future studies should be focused mainly on the assessment of such systems in free-living environments since in the study analysed there is a prevalence of controlled environments, and evaluating the reliability of smart insoles in continuous monitoring of daily life activities, or their application in clinical trials. Furthermore, the acceptance of the user of smart insole systems should be assessed, focusing on the comfort and usability aspects.

ACKNOWLEDGMENTS

This research is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 823978, and Invest Northern Ireland Proof of Concept project (PoC809). Luigi D'Arco is funded by the Ulster University Beitto Research Collaboration Programme.

REFERENCES

[1] Manuel Abbas, Dominique Somme, and Regine Le Bouquin Jeannes. 2021. D-SORM: A digital solution for remote monitoring based on the attitude of wearable devices. *Computer Methods and Programs in Biomedicine* 208 (2021), 106247.

[2] Ibrahim Almuteb, Rui Hua, and Ya Wang. 2022. Smart insoles review (2008-2021): Applications, potentials, and future. *Smart Health* 25 (9 2022). https://doi.org/10.1016/J.SMHL.2022.100301

[3] Geng An, Graham McCalmont, Philip Morrow, Mingjing Yang, Haiying Wang, Sally McClean, and Huiru Zheng. 2019. Feasibility study of a smart insole based AI Gait Assistant system. In *BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering*. 1–5.

[4] Y. Anand and P. P. Joby. 2019. Energy Expenditure Calculation with Physical Activity Recognition Using Genetic Algorithms. *Lecture Notes in Computational Vision and Biomechanics* 30 (2019), 525–530. https://doi.org/10.1007/978-3-030-00665-5_52

[5] Rafael De Pinho André, Pedro Henrique F.S. Diniz, and Hugo Fuks. 2017. Bottom-up investigation: Human Activity Recognition based on feet movement and posture information. *ACM International Conference Proceeding Series* Part F131931 (9 2017). https://doi.org/10.1145/3134230.3134240

[6] Paola Ariza Colpas, Enrico Vicario, Emiro De-La-Hoz-Franco, Marlon Pineres-Melo, Ana Oviedo-Carrascal, and Fulvio Patara. 2020. Unsupervised human activity recognition using the clustering approach: A review. *Sensors* 20, 9 (2020), 2702.

[7] Stacy J. Morris Bamberg, Ari Y. Benbasat, Donna Moxley Scarborough, David E. Krebs, and Joseph A. Paradiso. 2008. Gait Analysis Using a Shoe-Integrated Wireless Sensor System. *IEEE Transactions on Information Technology in Biomedicine* 12, 4 (2008), 413–423. https://doi.org/10.1109/TITB.2007.899493

[8] Avinandan Banerjee, Pawan Kumar Singh, and Ram Sarkar. 2021. Fuzzy Integral-Based CNN Classifier Fusion for 3D Skeleton Action Recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 31 (6 2021), 2206–2216. Issue 6. https://doi.org/10.1109/TCSVT.2020.3019293

[9] Marco Brighi, Annalisa Franco, and Dario Maio. 2021. ActivityExplorer: A semi-supervised approach to discover unknown activity classes in HAR systems. *Pattern Recognition Letters* 151 (11 2021), 340–347. https://doi.org/10.1016/J.PATREC.2021.08.029

[10] Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)* 46, 3 (2014), 1–33.

[11] Allah Bux, Plamen Angelov, and Zulfiqar Habib. 2017. Vision Based Human Activity Recognition: A Review. In *Advances in Computational Intelligence Systems*, Plamen Angelov, Alexander Gegov, Chrisina Jayne, and Qiang Shen (Eds.). Springer International Publishing, Cham, 341–371.

[12] Diliang Chen, Golnoush Asaeikheybari, Huan Chen, Wenyao Xu, and Ming-Chun Huang. 2021. Ubiquitous Fall Hazard Identification With Smart Insole. *IEEE Journal of Biomedical and Health Informatics* 25, 7 (2021), 2768–2776. https://doi.org/10.1109/JBHI.2020.3046701

[13] Diliang Chen, Yi Cai, Xiaoye Qian, Rahila Ansari, Wenyao Xu, Kuo-Chung Chu, and Ming-Chun Huang. 2019. Bring gait lab to everyday life: Gait analysis in terms of activities of daily living. *IEEE Internet of Things Journal* 7, 2 (2019), 1298–1312.

[14] Ian Cleland, Basel Kikhia, Chris Nugent, Andrey Boytsov, Josef Hallberg, Kåre Synnes, Sally McClean, and Dewar Finlay. 2013. Optimal Placement of Accelerometers for the Detection of Everyday Activities. *Sensors* 13, 7 (2013), 9183–9200. https://doi.org/10.3390/s130709183

[15] L Minh Dang, Kyungbok Min, Hanxiang Wang, Md Jalil Piran, Cheol Hee Lee, and Hyeonjoon Moon. 2020. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognition* 108 (2020), 107561.

[16] Suresh Dara and Priyanka Tumma. 2018. Feature Extraction By Using Deep Learning: A Survey. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 1795–1801. https://doi.org/10.1109/ICECA.2018.8474912

[17] Luigi D'Arco, Haiying Wang, and Huiru Zheng. 2022. Assessing Impact of Sensors and Feature Selection in Smart-Insole-Based Human Activity Recognition. *Methods and protocols* 5 (6 2022). Issue 3. https://doi.org/10.3390/MPS5030045

[18] Luigi D'Arco, Haiying Wang, and Huiru Zheng. 2022. A Rapid Detection of Parkinson's Disease using Smart Insoles: A Statistical and Machine Learning Approach. *Proceedings - 2022 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2022* (2022), 2985–2992. https://doi.org/10.1109/BIBM55620.2022.9995237

[19] Ruth M Davis. 1977. Evolution of computers and computing. *Science* 195, 4283 (1977), 1096–1102.

[20] Anis Davoudi, Mamoun T Mardini, David Nelson, Fahd Albinali, Sanjay Ranka, Parisa Rashidi, Todd M Manini, et al. 2021. The Effect of Sensor Placement and Number on Physical Activity Recognition and Energy Expenditure Estimation in Older Adults: Validation Study. *JMIR mHealth and uHealth* 9, 5 (2021), e23681.

[21] Fábio Carlos Lucas de Oliveira, Laurent Julien Bouyer, Amanda L. Ager, and Jean-Sébastien Roy. 2017. Electromyographic analysis of rotator cuff muscles in patients with rotator cuff tendinopathy: A systematic review. *Journal of Electromyography and Kinesiology* 35 (2017), 100–114. https://doi.org/10.1016/j.jelekin.2017.06.002

[22] Omid Dehzangi, Bhavani Anantapur Bache, and Omar Iftikhar. 2018. Activity Detection using Fusion of Multi-Pressure Sensors in Insoles. In *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 3315–3321. https://doi.org/10.1109/ICPR.2018.8545834

[23] Ton T.H. Duong, Leo Musacchia, David Uher, Jacqueline Montes, and Damiano Zanotto. 2022. Free-living Ambulatory Activity Classification: A Comparative Analysis of Wrist-worn, Insole-embedded, and Phone-embedded Sensors. *Proceedings of the IEEE RAS and*

EMBS International Conference on Biomedical Robotics and Biomechatronics 2022-August. https://doi.org/10.1109/BioRob52689.2022.9925432

[24] Luigi D'Arco, Haiying Wang, and Huiru Zheng. 2023. DeepHAR: a deep feed-forward neural network algorithm for smart insole-based human activity recognition. *Neural Computing and Applications 2023* (3 2023), 1–17. https://doi.org/10.1007/S00521-023-08363-W

[25] Changjun Fan and Fei Gao. 2021. Enhanced Human Activity Recognition Using Wearable Sensors via a Hybrid Feature Selection Method. *Sensors* 21, 19 (2021), 6434.

[26] Anna Ferrari, Daniela Micucci, Marco Mobilio, and Paolo Napoletano. 2021. Trends in human activity recognition using smartphones. *Journal of Reliable Intelligent Environments* 7 (9 2021), 189–213. Issue 3. https://doi.org/10.1007/S40860-021-00147-0/TABLES/7

[27] C. Ferri, J. Hernández-Orallo, and R. Modroiu. 2009. An experimental comparison of performance measures for classification. *Pattern Recognition Letters* 30, 1 (2009), 27–38. https://doi.org/10.1016/j.patrec.2008.08.010

[28] BJ Fogg. 2009. A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (Claremont, California, USA) *(Persuasive '09)*. Association for Computing Machinery, Article 40, 7 pages. https://doi.org/10.1145/1541948.1541999

[29] Wenbin Gao, Lei Zhang, Qi Teng, Jun He, and Hao Wu. 2021. DanHAR: dual attention network for multimodal human activity recognition using wearable sensors. *Applied Soft Computing* 111 (2021), 107728.

[30] Martin Gjoreski, Hristijan Gjoreski, Mitja Lutrek, and Matja Gams. 2015. Automatic Detection of Perceived Stress in Campus Students Using Smartphones. In *2015 International Conference on Intelligent Environments*. IEEE, 132–135. https://doi.org/10.1109/IE.2015.27

[31] Iván González, Jesús Fontecha, Ramón Hervás, and José Bravo. 2015. An ambulatory system for gait monitoring based on wireless sensorized insoles. *Sensors* 15, 7 (2015), 16589–16613.

[32] Rui Guo, Xiang Cheng, Zong-Chen Hou, Jing-Zhong Ma, Wen-Qiang Zheng, Xiao-Ming Wu, Dong Jiang, Yu Pan, and Tian-Ling Ren. 2021. A Shoe-Integrated Sensor System for Long-Term Center of Pressure Evaluation. *IEEE Sensors Journal* 21, 23 (2021), 27037–27044.

[33] Selda Güney and Çağatay Berke Erdaş. 2019. A Deep LSTM Approach for Activity Recognition. In *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. 294–297. https://doi.org/10.1109/TSP.2019.8768815

[34] Marian Haescher, Denys Matthies, Gerald Bieber, and Bodo Urban. 2015. CapWalk: A Capacitive Recognition of Walking-Based Activities as a Wearable Assistive Technology. *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (07 2015). https://doi.org/10.1145/2769493.2769500

[35] Nagaraj Hegde, Matthew Bries, Tracy Swibas, Edward Melanson, and Edward Sazonov. 2017. Automatic recognition of activities of daily living utilizing insole-based and wrist-worn wearable sensors. *IEEE journal of biomedical and health informatics* 22, 4 (2017), 979–988.

[36] Nagaraj Hegde, Edward Melanson, and Edward Sazonov. 2016. Development of a real time activity monitoring Android application utilizing SmartStep. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 1886–1889. https://doi.org/10.1109/EMBC.2016.7591089

[37] Nagaraj Hegde, Ting Zhang, Gitendra Uswatte, Edward Taub, Joydip Barman, Staci McKay, Andrea Taylor, David M Morris, Angi Griffin, and Edward S Sazonov. 2017. The pediatric SmartShoe: wearable sensor system for ambulatory monitoring of physical activity and gait. *IEEE transactions on neural systems and rehabilitation engineering* 26, 2 (2017), 477–486.

[38] Zawar Hussain, David Waterworth, Adnan Mahmood, Quan Z Sheng, and Wei Emma Zhang. 2021. Dataset for toothbrushing activity using brush-attached and wearable sensors. *Data in Brief* 37 (2021), 107248.

[39] Gu-Min Jeong, Phuc Huu Truong, and Sang-Il Choi. 2017. Classification of three types of walking activities regarding stairs using plantar pressure sensors. *IEEE Sensors Journal* 17, 9 (2017), 2638–2639.

[40] Kaleb Key, Spencer Mason, and Billy Kihei. 2021. Analysis of Activities based on a Single-Shoe Equipped with Force Sensors. *2021 IEEE International Smart Cities Conference, ISC2 2021* (9 2021). https://doi.org/10.1109/ISC253183.2021.9562856

[41] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. 2014. A survey of feature selection and feature extraction techniques in machine learning. In *2014 Science and Information Conference*. IEEE, 372–378. https://doi.org/10.1109/SAI.2014.6918213

[42] Sabri Khssibi, Hanen Idoudi, Adrien Van Den Bossche, Thierry Val, and Leila Azzouz Saidane. 2013. Presentation and analysis of a new technology for low-power wireless sensor network. *International Journal of Digital Information and Wireless Communications* 3, 1 (2013), pp–75.

[43] Sangyeon Kim, Gawon Lee, and Jihie Kim. 2020. Lightweight Real-time Fall Detection using Bidirectional Recurrent Neural Network. In *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*. IEEE, 1–5. https://doi.org/10.1109/SCISISIS50064.2020.9322735

[44] Leanne M Kmet, Linda S Cook, and Robert C Lee. 2004. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. *Edmonton: Alberta Heritage Foundation for Medical Research (AHFMR)* (2004). https://doi.org/10.7939/R37M04F16

[45] Kyoungchul Kong and Masayoshi Tomizuka. 2009. A Gait Monitoring System Based on Air Pressure Sensors Embedded in a Shoe. *IEEE/ASME Transactions on Mechatronics* 14, 3 (2009), 358–370. https://doi.org/10.1109/TMECH.2008.2008803

[46] Vipin Kumar and Sonajharia Minz. 2014. Feature selection: a literature review. *SmartCR* 4, 3 (2014), 211–229.

[47] Guohao Lan, Dong Ma, Weitao Xu, Mahbub Hassan, and Wen Hu. 2020. Capacitor-Based Activity Sensing for Kinetic-Powered Wearable IoTs. *ACM Trans. Internet Things* 1, 1, Article 2 (March 2020), 26 pages. https://doi.org/10.1145/3362124

[48] Oscar D Lara and Miguel A Labrador. 2012. A survey on human activity recognition using wearable sensors. *IEEE communications surveys & tutorials* 15, 3 (2012), 1192–1209.

[49] Marco Leo, Tiziana D'Orazio, and Paolo Spagnolo. 2004. Human Activity Recognition for Automatic Visual Surveillance of Wide Areas. In *Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks* (New York, NY, USA) *(VSSN '04)*. Association for Computing Machinery, New York, NY, USA, 124–130. https://doi.org/10.1145/1026799.1026820

[50] Feng Lin, Aosen Wang, Yan Zhuang, Machiko R. Tomita, and Wenyao Xu. 2016. Smart Insole: A Wearable Sensor Device for Unobtrusive Gait Monitoring in Daily Life. *IEEE Transactions on Industrial Informatics* 12, 6 (2016), 2281–2291. https://doi.org/10.1109/TII.2016.2585643

[51] Zhirong Lin, Zengwei Wang, Houde Dai, and Xuke Xia. 2022. Efficient fall detection in four directions based on smart insoles and RDAE-LSTM model. *Expert Systems with Applications* 205 (11 2022), 117661. https://doi.org/10.1016/J.ESWA.2022.117661

[52] Kai-Chun Liu, Chia-Yeh Hsieh, Steen Jun-Ping Hsu, and Chia-Tai Chan. 2018. Impact of Sampling Rate on Wearable-Based Fall Detection Systems Based on Machine Learning Models. *IEEE Sensors Journal* 18, 23 (2018), 9882–9890. https://doi.org/10.1109/JSEN.2018.2872835

[53] Hugo Luís, Luís Garrote, and Urbano J. Nunes. 2021. Human Activity Recognition for Indoor Robotics: A Deep Learning Based Approach Using a Human Detection Stage. In *2021 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. 43–49. https://doi.org/10.1109/ICARSC52212.2021.9429774

[54] Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de las Heras. 2019. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition* 91 (2019), 216–231.

[55] Ojan Majidzadeh Gorjani, Radek Byrtus, Jakub Dohnal, Petr Bilik, Jiri Koziorek, and Radek Martinek. 2021. Human Activity Classification Using Multilayer Perceptron. *Sensors* 21, 18 (2021), 6207.

[56] Denys J.C. Matthies, Thijs Roumen, Arjan Kuijper, and Bodo Urban. 2017. CapSoles: Who is walking on what kind of floor? *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2017.* https://doi.org/10.1145/3098279.3098545

[57] Graham McCalmont, Philip Morrow, Huiru Zheng, Anas Samara, Sara Yasaei, Haiying Wang, and Sally McClean. 2018. eZiGait: Toward an AI Gait Analysis And Sssistant System. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2280–2286. https://doi.org/10.1109/BIBM.2018.8621176

[58] Sakorn Mekruksavanich and Anuchit Jitpattanakul. 2022. Multimodal Wearable Sensing for Sport-Related Activity Recognition Using Deep Learning Networks. *Journal of Advances in Information Technology* 13 (2022). Issue 2. https://doi.org/10.12720/JAIT.13.2.132-138

[59] Kohle J Merry, Evan Macdonald, Megan MacPherson, Omar Aziz, Edward Park, Michael Ryan, and Carolyn J Sparrey. 2021. Classifying sitting, standing, and walking using plantar force data. *Medical & Biological Engineering & Computing* 59, 1 (2021), 257–270.

[60] Michelle E. Mlinac and Michelle C. Feng. 2016. Assessment of Activities of Daily Living, Self-Care, and Independence. *Archives of Clinical Neuropsychology* 31, 6 (08 2016), 506–516. https://doi.org/10.1093/arclin/acw049

[61] Christopher Moufawad el Achkar, Constanze Lenoble-Hoskovec, Anisoara Paraschiv-Ionescu, Kristof Major, Christophe Büla, and Kamiar Aminian. 2016. Instrumented shoes for activity classification in the elderly. *Gait & Posture* 44 (2016), 12–17. https://doi.org/10.1016/j.gaitpost.2015.10.016

[62] Christopher Moufawad el Achkar, Constanze Lenoble-Hoskovec, Anisoara Paraschiv-Ionescu, Kristof Major, Christophe Büla, and Kamiar Aminian. 2016. Physical Behavior in Older Persons during Daily Life: Insights from Instrumented Shoes. *Sensors* 16, 8 (2016). https://doi.org/10.3390/s16081225

[63] Armelle M. Ngueleu, Andréanne K. Blanchette, Désirée Maltais, Hélène Moffet, Bradford J. McFadyen, Laurent Bouyer, and Charles S. Batcho. 2019. Validity of Instrumented Insoles for Step Counting, Posture and Activity Recognition: A Systematic Review. *Sensors 2019, Vol. 19, Page 2438* 19 (5 2019), 2438. Issue 11. https://doi.org/10.3390/S19112438

[64] Nhan Duc Nguyen, Duong Trong Bui, Phuc Huu Truong, and Gu-Min Jeong. 2018. Classification of five ambulatory activities regarding stair and incline walking using smart shoes. *IEEE Sensors Journal* 18, 13 (2018), 5422–5428.

[65] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). https://doi.org/10.1136/bmj.n71

[66] Abhishesh Pal, Antonio C. Leite, Jon G. O. Gjevestad, and Pal J. From. 2021. A Video-Based Human Activity and Motion Direction Classification Framework for Agricultural Fields. In *2021 6th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*. IEEE, 1–7. https://doi.org/10.1109/ACIRS52449.2021.9519337

[67] I.P.I. Pappas, M.R. Popovic, T. Keller, V. Dietz, and M. Morari. 2001. A reliable gait phase detection system. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 9, 2 (2001), 113–125. https://doi.org/10.1109/7333.928571

[68] Ashish Patel and Jigarkumar Shah. 2020. Real-time human behaviour monitoring using hybrid ambient assisted living framework. *Journal of Reliable Intelligent Environments* 6, 2 (2020), 95–106.

[69] Arman J. Paydarfar, Antonio Prado, and Sunil K. Agrawal. 2020. Human Activity Recognition Using Recurrent Neural Network Classifiers on Raw Signals from Insole Piezoresistors. In *2020 8th IEEE RAS/EMBS International Conference for Biomedical Robotics and*

Biomechatronics (BioRob). IEEE, 916–921. https://doi.org/10.1109/BioRob49111.2020.9224311

[70] Cuong Pham, Linh Nguyen, Anh Nguyen, Ngon Nguyen, and Van Toi Nguyen. 2021. Combining skeleton and accelerometer data for human fine-grained activity recognition and abnormal behaviour detection with deep temporal convolutional networks. *Multimedia Tools and Applications 2021 80:19* 80 (6 2021), 28919–28940. Issue 19. https://doi.org/10.1007/S11042-021-11058-W

[71] Bronagh Quigley, Mark Donnelly, George Moore, and Leo Galway. 2018. A Comparative Analysis of Windowing Approaches in Dense Sensing Environments. *Proceedings* 2, 19 (2018). https://doi.org/10.3390/proceedings2191245

[72] Dian Ren, Nathanael Aubert-Kato, Emi Anzai, Yuji Ohta, and Julien Tripette. 2020. Random forest algorithms for recognizing daily life activities using plantar pressure information: a smart-shoe study. *PeerJ* 8 (2020), e10170.

[73] Salma Saidani, Rim Haddad, Ridha Bouallegue, and Raed Shubair. 2022. Smart Insole Monitoring System for Fall Detection and Bad Plantar Pressure. *Lecture Notes in Networks and Systems* 451 LNNS (2022), 199–208. https://doi.org/10.1007/978-3-030-99619-2_20/COVER

[74] Nilotpola Sarma, Suchetana Chakraborty, and Dip Sankar Banerjee. 2019. Learning and Annotating Activities for Home Automation using LSTM. *2019 11th International Conference on Communication Systems and Networks, COMSNETS 2019* (5 2019), 631–636. https://doi.org/10.1109/COMSNETS.2019.8711433

[75] Edward Sazonov, Nagaraj Hegde, Raymond C Browning, Edward L Melanson, and Nadezhda A Sazonova. 2015. Posture and activity recognition and energy expenditure estimation in a wearable platform. *IEEE journal of biomedical and health informatics* 19, 4 (2015), 1339–1346.

[76] Kimiaki Shirahama and Marcin Grzegorzek. 2017. On the generality of codebook approach for sensor-based human activity recognition. *Electronics* 6, 2 (2017), 44.

[77] Muhammad Shoaib, Hans Scholten, and P.J.M. Havinga. 2013. Towards Physical Activity Recognition Using Smartphone Sensors. In *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*. IEEE, 80–87. https://doi.org/10.1109/UIC-ATC.2013.43

[78] Deepika Singh, Ismini Psychoula, Johannes Kropf, Sten Hanke, and Andreas Holzinger. 2018. Users' perceptions and attitudes towards smart home technologies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10898 LNCS (2018), 203–214. https://doi.org/10.1007/978-3-319-94523-1_18/COVER/

[79] Ahmed Snoun, Nozha Jlidi, Tahani Bouchrika, Olfa Jemai, and Mourad Zaied. 2021. Towards a deep human activity recognition approach based on video to image transformation with skeleton data. *Multimedia Tools and Applications 2021 80:19* 80 (7 2021), 29675–29698. Issue 19. https://doi.org/10.1007/S11042-021-11188-1

[80] Liwen Song, Changcheng Xiang, Huafeng Guo, and Shiqiang Chen. 2021. Activity recognition using multiplex limited penetrable visibility graph. *Journal of the mechanical behavior of biomedical materials* 124 (12 2021). https://doi.org/10.1016/J.JMBBM.2021.104880

[81] Stephen V Stehman. 1997. Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment* 62, 1 (1997), 77–89.

[82] Sophini Subramaniam, Sumit Majumder, Abu Ilius Faisal, and M. Jamal Deen. 2022. Insole-Based Systems for Health Monitoring: Current Solutions and Research Challenges. *Sensors* 22, 2 (2022). https://doi.org/10.3390/s22020438

[83] Phuc Truong, Sujeong You, Ji Sang Hoon, and Gu-Min Jeong. 2019. Wearable System for Daily Activity Recognition Using Inertial and Pressure Sensors of a Smart Band and Smart Shoes. *International Journal of Computers Communications & Control* 14 (11 2019), 726–742. https://doi.org/10.15837/ijccc.2019.6.3618

[84] Phuc Huu Truong, Sujeong You, Sang Hoon Ji, and Gu Min Jeong. 2021. Adaptive Accumulation of Plantar Pressure for Ambulatory Activity Recognition and Pedestrian Identification. *Sensors 2021, Vol. 21, Page 3842* 21 (6 2021), 3842. Issue 11. https://doi.org/10.3390/S21113842

[85] Lan Wang, Min Peng, and Qing F. Zhou. 2019. Fall Detection Based on Convolutional Neural Networks Using Smart Insole. In *2019 5th International Conference on Control, Automation and Robotics (ICCAR)*. IEEE, 593–598. https://doi.org/10.1109/ICCAR.2019.8813332

[86] Zhen Wei, Zhiwang Zhang, Jiayi Jiang, Yu Zhang, and Lin Wang. 2019. Comparison of plantar loads among runners with different strike patterns. *https://doi.org/10.1080/02640414.2019.1623990* 37 (9 2019), 2152–2158. Issue 18. https://doi.org/10.1080/02640414.2019.1623990

[87] Mark Weiser. 1999. The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.* 3, 3 (jul 1999), 3–11. https://doi.org/10.1145/329124.329126

[88] Kieran Woodward, Eiman Kanjo, Andreas Oikonomou, and Alan Chamberlain. 2020. LabelSens: enabling real-time sensor data labelling at the point of collection using an artificial intelligence-based approach. *Personal and Ubiquitous Computing* 24 (10 2020), 709–722. Issue 5. https://doi.org/10.1007/S00779-020-01427-X/FIGURES/10

[89] Cunyi Yin, Jing Chen, Xiren Miao, Hao Jiang, and Deying Chen. 2021. Device-Free Human Activity Recognition with Low-Resolution Infrared Array Sensor Using Long Short-Term Memory Neural Network. *Sensors* 21, 10 (2021), 3551.

[90] Kuan Zhang, Ming Sun, D. Kevin Lester, F. Xavier Pi-Sunyer, Carol N. Boozer, and Richard W. Longman. 2005. Assessment of human locomotion by using an insole measurement system and artificial neural networks. *Journal of Biomechanics* 38, 11 (2005), 2276–2287. https://doi.org/10.1016/j.jbiomech.2004.07.036

[91] Xiaotong Zhang and Jin Zhang. 2019. Subject Independent Human Activity Recognition with Foot IMU Data. In *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*. IEEE, 240–246. https://doi.org/10.1109/MSN48538.2019.00054

[92] Ömer Faruk, Ertu˘grul Ertu˘grul, and Yılmaz Kaya. 2017. Determining the optimal number of body-worn sensors for human activity recognition. *Soft Computing* 21 (2017), 5053–5060. https://doi.org/10.1007/s00500-016-2100-7

A ACRONYMS

3D-ACC Tri-dimension Accelerometer.
3D-GYR Tri-dimension Gyroscope.
3D-MAG Tri-dimension Magnetometer.

ADLs Activities of Daily Living.
ANN Artificial Neural Network.
ANOVA Analysis of Variance.
AR Auto Regressive Parameters.
AUC Area Under the Curve.

BAR Barometer.
BLE Bluetooth Low Energy.
BW Body Weight.

CNN Convolutional Neural Network.
CP Cerebral Palsy Participants.

DA Discriminant Analysis.
DT Decision tree.

ECU Electronic Control Unit.

F Female.
FN False Negative.
FP False Positive.
FSR Force Sensor Resistor.

GNB Gaussian Naive Bayes.
GQ Good Quality.
GRF Ground Reaction Force.

HAR Human Activity Recognition.
HQ High Quality.

IMU Inertial Measurement Unit.

kF-CV k-fold Cross-Validation.
KNN K-Nearest Neighbours.

LOO-CV Leave-One-Out Cross-Validation.
LOSO-CV Leave-One-Subject-Out Cross-Validation.
LQ Low Quality.

M Male.

MLD Multinomial Logistic Discrimination.
MLP Multi-Layer Perceptron.
MLPVG Multi-layer Limited Penetrable Visibility Graph.
MQ Moderate Quality.
MRMR Minimum Redundancy Maximum Relevance.

N/A Not Applicable.
NB Naive Bayes.

PCA Principal Component Analysis.
PEH Piezoelectric Energy Harvester.
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
PS Pressure Sensor.
PSA Pressure Sensor Array.

RF Random Forest.
RNN Recurrent Neural Network.
ROC Receiver Operating Characteristic.

STFT Short Time Fourier Transform.
SVM Support Vector Machine.

TAWA Temporally Adaptive Weighting Accumulation.
TN True Negative.
TP True Positive.

USB Universal Serial Bus.