Fall detection using accelerometer-based smartphones: Where do we go from here?

Tristan Stampfler[1†], Mohamed Elgendi[1*†],
Richard Ribon Fletcher[2] and Carlo Menon[1]

[1]Biomedical and Mobile Health Technology Laboratory, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland, [2]Mobile Technology Group, Department of Mechanical Engineering, MIT, Cambridge, MA, United States

According to World Health Organization statistics, falls are the second leading cause of unintentional injury deaths worldwide. With older people being particularly vulnerable, detecting, and reporting falls have been the focus of numerous health technology studies. We screened 267 studies and selected 15 that detailed pervasive fall detection and alerting apps that used smartphone accelerometers. The fall datasets used for the analyses included between 4 and 38 participants and contained data from young and old subjects, with the recorded falls performed exclusively by young subjects. Threshold-based detection was implemented in six cases, while machine learning approaches were implemented in the other nine, including decision trees, k-nearest neighbors, boosting, and neural networks. All methods could ultimately achieve real-time detection, with reported sensitivities ranging from 60.4 to 99.3% and specificities from 74.6 to 100.0%. However, the studies had limitations in their experimental set-ups or considered a restricted scope of daily activities—not always representative of daily life—with which to define falls during the development of their algorithms. Finally, the studies omitted some aspects of data science methodology, such as proper test sets for results evaluation, putting into question whether reported results would correspond to real-world performance. The two primary outcomes of our review are: a ranking of selected articles based on bias risk and a set of 12 impactful and actionable recommendations for future work in fall detection.

Introduction

Falls are the second leading cause of unintentional injury deaths worldwide, according to the World Health Organization (1). More than 37 million falls each year are severe enough to require medical attention (1), and people over 60 years old are the most concerned with fatal falls (2). The global population of over-60s was estimated to be 962 million in 2017, a number which is expected to double by 2050 (3).

Furthermore, this population increasingly lives in isolated conditions, making falls even more dangerous because immediate assistance cannot always be provided (2). Other populations concerned with falls are younger people with disabilities or people recovering from an operation or an injury. There is thus a need for technologies to better handle falls, such as falling detection systems that can automatically send alerts. The core principle of phone-based fall detection is illustrated in Figure 1.

We offer here a review of research studies (4–18) that propose practical methods to detect falls using a pervasive device that most people carry every day: smartphones. These devices have many built-in sensors, including an accelerometer, which can capture the acceleration caused by a fall. Data can be collected continuously during the day, which can then be analyzed using threshold-based algorithms or more modern machine learning techniques. This raises the possibility of real-time operation, in which the chosen methodology would be expected to achieve real-time detection and alerting. The selected articles each report the building of an app that assumes a phone being kept in a common location, like in a pocket. The development of such fall apps presents four main challenges: to design an experimental set-up and build an appropriate dataset for fall recognition; to conceive and tune an algorithm; to evaluate the results, including defining relevant metrics and an evaluation procedure; and to ensure that the detection and all associated functionalities work well-together.

The choice of using a phone to carry such monitoring is part of the broader concept of "digital phenotyping," which is about tracking people's behavior using data from personal digital devices (19). Smartphones bring together virtually all characteristics one can wish to have for an efficient monitoring: a huge adoption rate, a multitude of sensors, computing power and network connection. These make them especially relevant for fall detection but also other tasks such as general gait analysis or mental health assessment. Van Laerhoven et al. (20) determined that users have their smartphones on them on average 36% of daytime. The main pattern they describe is that while sitting, users tend to put their phone on the table and place it in their pockets when they move, which suits well a phone-based fall detection paradigm discussed here. The trousers' pocket emplacement is the first to carry a smartphone when on the body for both men (21) and women (22). However, there is also a significant gender difference as carrying a phone separately in a bag is also very popular among women (22) and only mildly adopted among men (21).

This review aims to provide a detailed overview of existing solutions reported as working for fall detection using smartphone accelerometers. We describe the various approaches, highlight best practices, and note areas for improvement, presenting a review that will give valuable insights for future research into fall detection and offer a significant head-start for future development.

Related work and technologies

First we would like to shortly mention another review work from Casilari et al. (23), its main conclusions but also limitations regarding fall detection with a smartphone. The study presents quite extensively the technological and IT possibilities for a fall detection system. However from our perspective, the most relevant insights from this study lie on its critique on evaluation methods being performed to assess a solution performance. It notably points out the lack of common data across studies to draw comparisons, the heterogeneity of fall typologies used and

FIGURE 1
Principle for fall detection with a smartphone, our review focuses on the detection step.
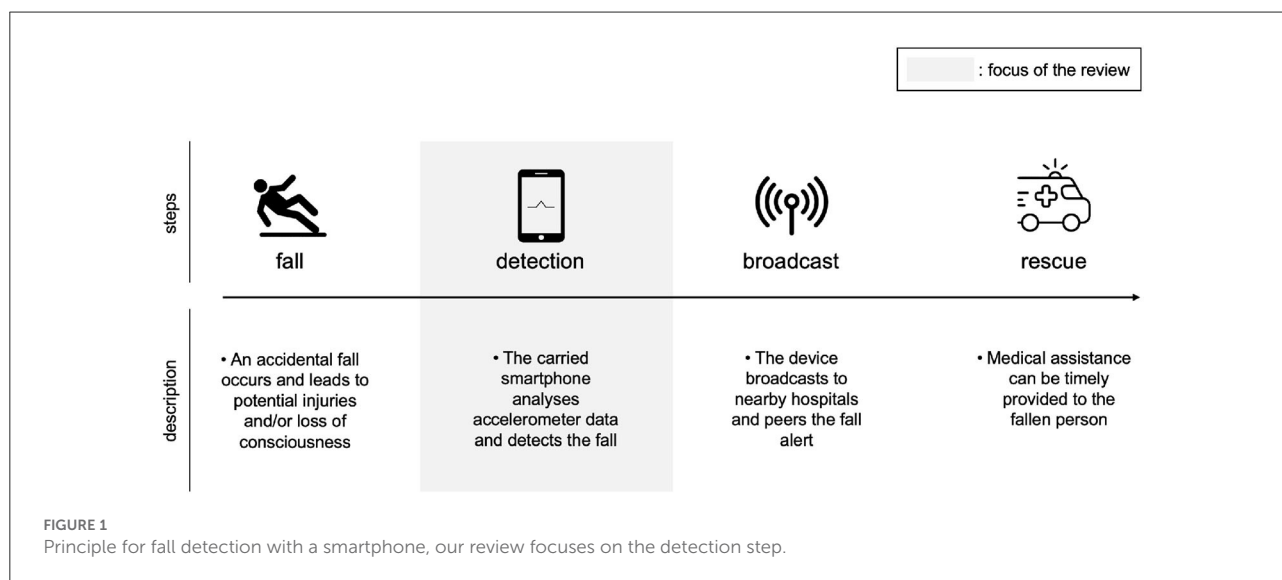
the lack of real world evaluation. Nevertheless, this review clearly has non negligible limitations. From a timeline perspective it was published in 2015 which means it could not capture most recent papers and development of the topic. This is particularly illustrated by a lack of discussion on machine learning choices and issues one must face when developing a time series classifier for fall detection.

Second, we also would like to indicate alternative technologies that have been explored to detect falls beyond using the smartphone. It is possible to group them into two separate categories: wearable sensors and external sensors. The first category entails wearing on the body a sensor (commonly an accelerometer), which is usually built in a smartwatch or a belt. Some commercial application have even seen the day of light with for instance the Apple Watch offering a fall detection feature to its users. The second category corresponds generally to devices that can be found in a smart-home and that can monitor their environment. The list includes radars, infrared sensors, ultrasounds, or even image processing of camera recordings. For further reference on technological possibilities for fall detection, we recommend the review from Wang et al. (24).

Materials and methods

Search strategy

To conduct our review, we searched three research databases-*IEEE Xplore*, *PubMed*, and *Embase*-chosen for their relevance to this topic at the nexus of health and modern data analysis. We used their advanced keyword search functionalities to find articles published in the last decade (from January 1st, 2012, to the date of our search on March 17th, 2022). Our search request contained the following search terms: (fall OR falling) AND (accelerometer) AND [smartphone OR (smart phone)].

Inclusion and exclusion criteria

To refine the initial search results, we applied the following additional criteria based on content. **Screening**: the article must be readable by the reviewers. **Type**: the article should be a primary research study. **Goal**: the study should be focused on fall detection. **Device**: the employed device should be a smartphone with an accelerometer. **Phone positioning**: the phone should be positioned in a natural and common way without additional straps. **App**: the research should have led to the development of an app as a proof-of-concept.

Results

Applying these criteria as filters, we reduced our initial results set of 267 articles to the 15 most relevant to understand the current state of the art. The complete process can be visualized with the PRISMA diagram in Figure 2.

In this section, we aim to provide a comprehensive comparative view of the selected articles by analyzing them according to a set of features and metrics. Table 1 summarizes all information collected on the studies. After a careful reading of each article, we collated their content across the following dimensions:
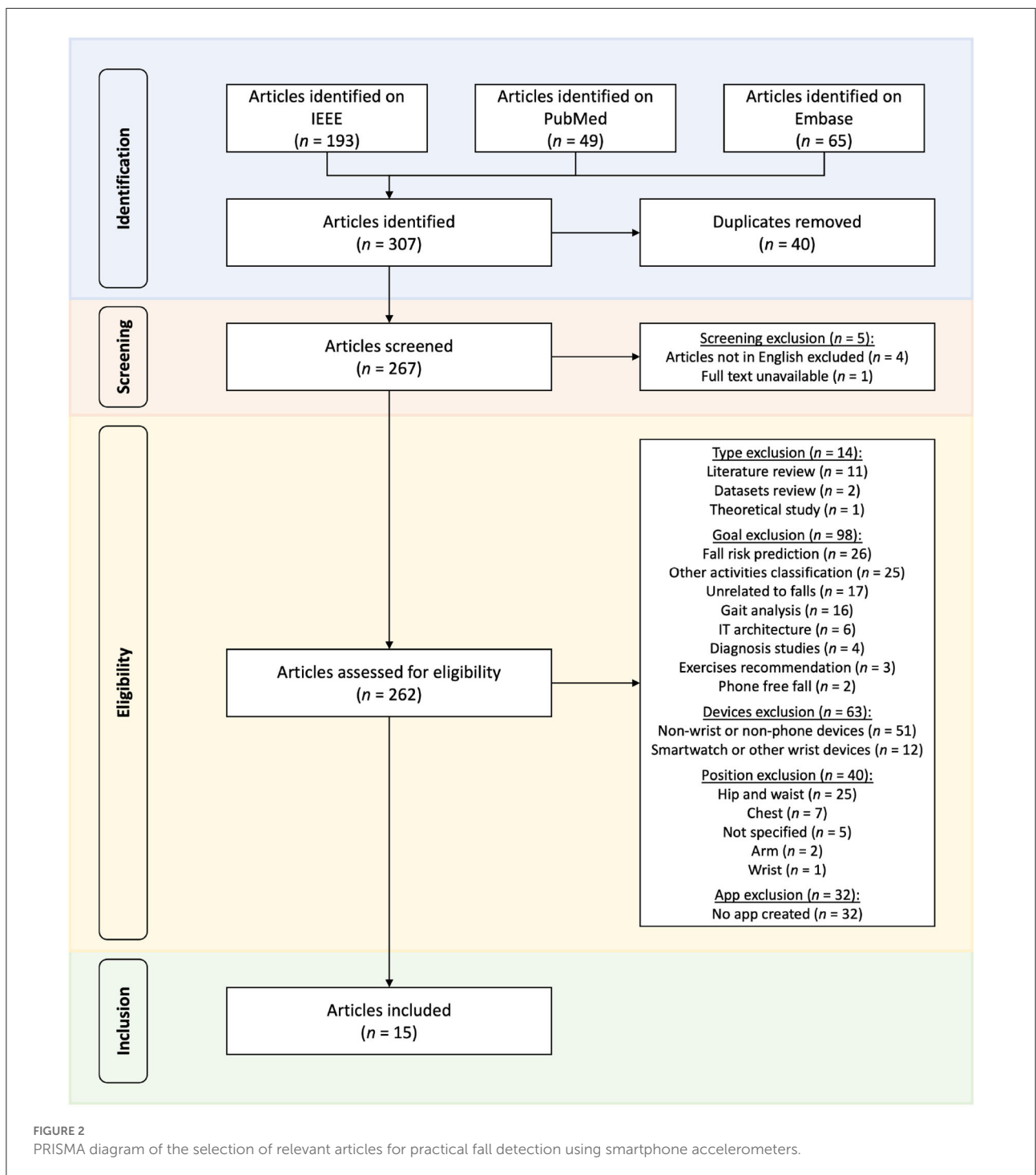
- **Publication:** general information on the publication to provide context
- **Data collection:** characteristics of experimental set-up characteristics to understand how the data were collected
- **Scope:** the range and the depth of the investigation
- **Data analysis:** techniques used to illustrate the design of the proposed algorithms
- **Evaluation:** metrics reported to compare performances

Publication

The selected studies span most of the last decade, ranging from 2012 to 2020. The average and median publication year is 2016, which is in contrast with the concurrent rapid development of data science. It should therefore be kept in mind that earlier studies (4–10) did not have access to same knowledge and tools for their analysis as the most recent ones (13–18). A visualization of the publication years is presented in Figure 3 and highlights a steady interest for the topic.

Data collection

For data collection, most of the studies (4–12, 14–17) designed their own experimental set-ups and constructed their own fall datasets from scratch. The numbers of participants were relatively low, with an average of 14.8, but this was mostly due to the laborious experimental procedures. Figure 4 displays the age ranges of the participants as well as their reported health conditions. We note in particular that two studies (5, 18) included real data from elderly people, who generally performed only the activities of daily living (ADLs), but not falls. No study mentions having people with disabilities or recovering from an operation in their data. One study (9) had the ingenious idea of adding a dummy with human proportions and weights as a means to simulate difficult falls.

FIGURE 2
PRISMA diagram of the selection of relevant articles for practical fall detection using smartphone accelerometers.

All the selected studies considered the common positioning of phones in pants pockets, although some also investigated a second positioning. In some cases, the researchers were looking for better performance by placing the phone at a higher, yet less common, position on the body, such as the waist (18), belt (10, 14), or shirt pocket (7). In other cases, the researchers focused on additional real-life possibilities, like having the phone in the hand (8, 16) or in a bag (6). Because of the exclusion criteria, all the studies used the accelerometer, with only two (7, 16) leveraging angle sensors

TABLE 1 Summary table of the 15 selected articles.

Publication									Data collection				Data analysis			Evaluation				
References	Year	No. ADLs	No. Falls	Phone drop	Data-set[a]	No. Part	Age	Gender (M,F)	Device position[b]	Sensors[c]	Sampl. freq.	Falls reps	Main features[d]	Filters[e]	Algorithm	Sensi-tivity	Speci-ficity	Accu-racy	Eval. method	Power check
Shi et al. (4)	2012	5	5	Yes	O	4	20–26	3, 1	TP	A	50 Hz	100	GR, MA, TI	LP	SVM	90.0	95.7	N/A	LOSO	No
Aguiar et al. (5)	2014	10	10	No	O	36	22–69	28, 8	TP	A	67 Hz	1,025	SF,MA	LP	Dec. Tree	94.0	90.2	92.0	CV	Yes
Medrano et al. (6)	2014	N/A	8	No	O	10	20–42	7, 3	TP,BA	A	50 Hz	240	N/A	N/A	kNN	92.0	95.0	N/A	CV	No
Pierleoni et al. (7)	2015	10	10	No	O	5	24–29	N/A	TP,SP	A,M	40 Hz	150	SF	LP	SVM	99.3	96.0	97.7	Train set	No
Helmy and Helmy (8)	2015	N/A	4	Yes	O	14	N/A	N/A	TP,OH	A	20 Hz	400	TI,MA	N/A	Thresholds	95.0	90.0	N/A	Train set	Yes
Cao et al. (9)	2015	N/A	17	No	O	20	N/A	N/A	TP	A	17 Hz	648	MA	N/A	Boosting	88.0	74.6	N/A	Train set	Yes
Vermeulen et al. (10)	2015	16	10	No	O	8	18–24	4, 4	TP,BE	A	67 Hz	400	MA	N/A	Thresholds	90.0	87.0	N/A	Train set	No
Srisuphab et al. (11)	2016	N/A	9	No	O	N/A	N/A	N/A	TP	A	N/A	90	TI,MA,VE	N/A	Thresholds	95.6	N/A	N/A	Train set	No
Chaitep and Chawachat (12)	2017	2	8	Yes	O	N/A	N/A	N/A	TP	A	N/A	800	MA,GF	N/A	Thresholds	83.0	91.5	N/A	Train set	No
Tsinganos and Skodras (13)	2017	12	4	No	P	24	22–47	17, 7	TP	A	50 Hz	288	TD,SF	LP	kNN	97.5	94.9	N/A	CV	Yes
Tran et al. (14)	2017	3	1	No	O	15	N/A	N/A	TP,BE	A	50 Hz	86	MA,TD	N/A	Perceptron	60.4	94.8	82.5	Val. set	No
Shahzad and Kim (15)	2018	10	10	No	O	4	26–32	3, 1	TP	A	64 Hz	175	TI,MA	LP	MKL-SVM	95.8	88.0	91.7	CV	Yes
Ning et al. (16)	2018	5	4	No	O	10	N/A	6, 4	TP,OH	A,G,M	50 Hz	400	MA	KA	Thresholds	92.0	91.2	90.0	Train set	No
Lee and Tseng (17)	2019	11	4	No	O	4	N/A	4, 0	TP	A	50 Hz	100	MA	N/A	Thresholds	96.0	100	99.0	Train set	No
Salama and Eskaf (18)	2020	19	15	No	P	38	19–75	19, 19	TP,WA	A	50 Hz	1,798	SF,MA	ME	DNN	N/A	N/A	96.1	Test set	Yes

[a]O, Own; P, Public.
[b]TP, Trouser's pocket; SP, Shirt's pocket; WA, Waist; BE, Belt; OH, On-hand.
[c]A, Accelerometer; M, Magnetometer; G, Gyroscope.
[d]GR, gravity cross-rate; MA, magnitude; TI, tilt; SF, statistical features; VE, velocity; GF, G-force; TD, time domain.
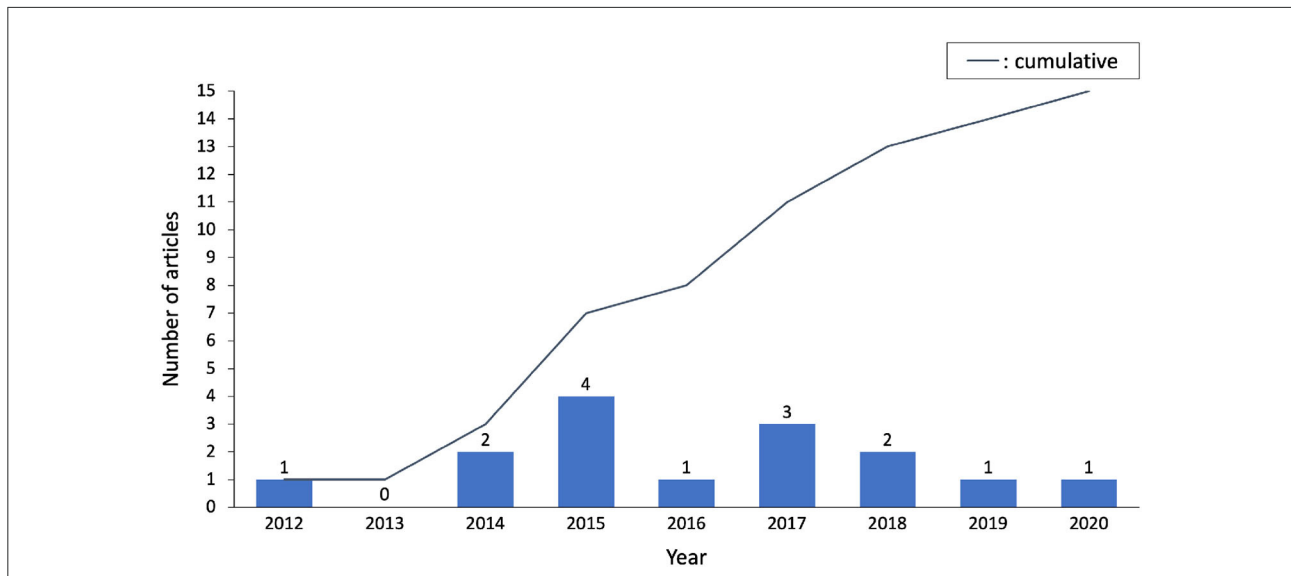[e]LP, low-pass; KA, Kalman; ME, mean.

FIGURE 3
Publication years of selected papers, there is a steady interest for the topic of app-based fall detection.
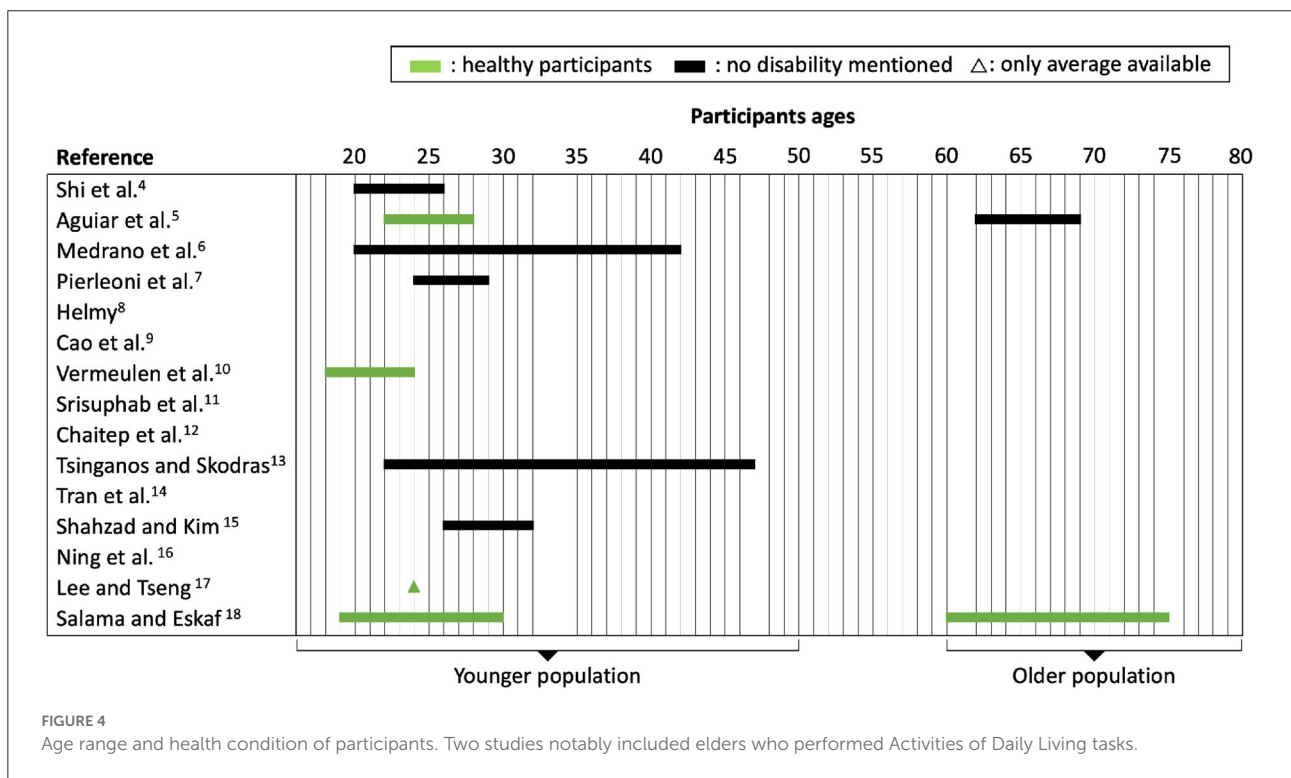
FIGURE 4
Age range and health condition of participants. Two studies notably included elders who performed Activities of Daily Living tasks.

as well. Most sampling frequencies (4, 6–8, 10, 13–18) were in the range of 40–60 Hz, with three studies (7, 14, 18) interestingly mentioning that this represents a good trade-off between precision and efficiency. Clearly, with a higher sampling frequency, more data points are collected, but more computational power is then required for both sampling and

analysis. One of the articles (9) describes the frequency as a trade-off in precision itself, with an excessively high frequency creating noise and one that is too low missing important signals. Finally, Nyquist's theorem is also mentioned (13), according to which the sampling frequency should be at least double that of the typical 0–20 Hz frequency of human movements.

FIGURE 5
Our plot of acceleration magnitudes for different types of activities of daily living and falls, based on data retrieved from the *UniMiB-SHAR* (25) dataset. Amplitude and signal frequency differences are apparent from visual inspection.
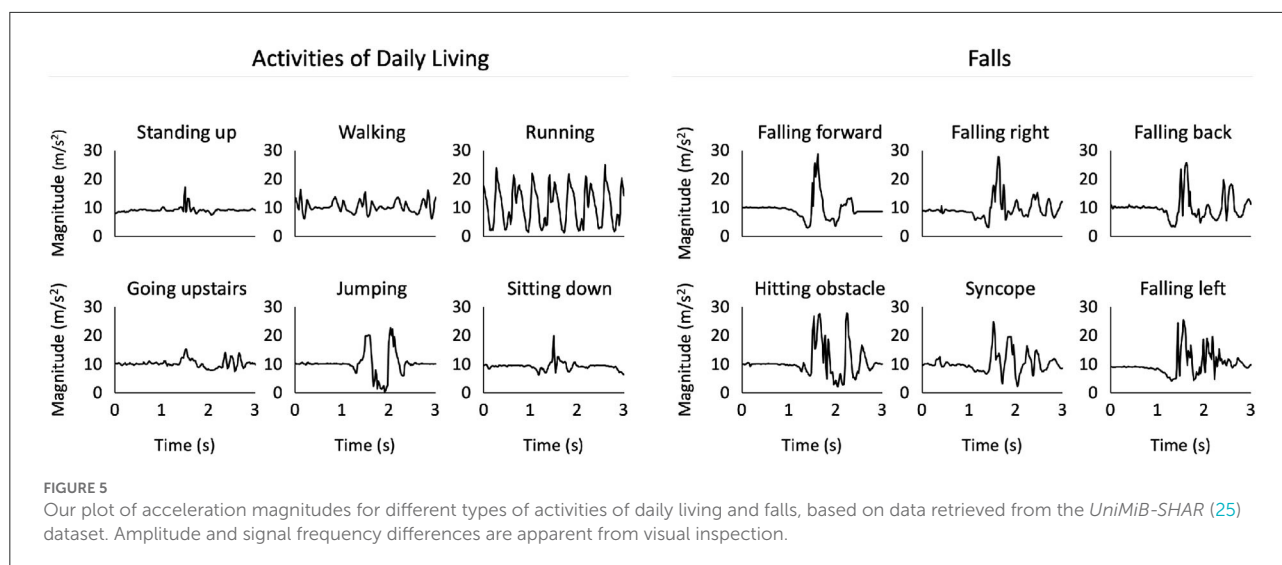
To give the reader a better idea of the signals being recorded, we plotted the acceleration magnitude of ADL and fall samples from the public *UniMiB-SHAR* (25) dataset (Figure 5). The plot is divided into two groups: ADLs and falls. There are clear differences in terms of signal magnitude and frequencies between the two groups. Even inside the ADLs group, distinctive patterns appear based on the activity being performed. Whilst walking and running feature periodic signals, other activities like standing up or jumping lead to a more sudden, one-time acceleration signal. Regarding falls, they all share a sudden rise in acceleration magnitude when the fall occurs and seem to have similar patterns. It is however to keep in mind that the accelerometer records separately the acceleration across the three movement axis (not shown here). Such separation can help classification algorithms distinguish further sub categories of falls.

Scope

The scope represents the extent to which fall detection was studied. Eleven articles (4, 5, 7, 10, 12–18) reportedly used pre-defined ADLs as negative samples, including activities like walking, sitting, jumping, and walking up and downstairs. While this makes sense from a classification perspective, pre-defining ADLs creates the risk of omitting real-life situations that should be considered to be non-falls. Two articles (6, 9) took the approach of leaving subjects to their normal daily routines and passively recording data while they carried their phones. This latter approach has the advantage of being unbiased regarding the types of daily activities, though it may still miss some rare non-fall events.

Falls themselves were simulated in all cases because recording real unwanted falls would be very challenging due to their rarity. Different fall types were envisioned in all but one study (4–13, 15–18), even though the most dangerous scenarios (like falling down stairs) were never performed, which puts the practicality of the proposed solutions into question as they were not conceived to detect the most perilous types of falls. Finally, the possibility of dropping the phone was only discussed and addressed in three instances (4, 8, 12) despite it being a key event to consider due to its ordinariness and its fall-like pattern.

Data analysis

Feature engineering is often performed after data are gathered from sensors, but before classification. All but one study (4, 5, 7–18) used orientation-invariant features, like magnitude-such orientation-independent methods are preferable to ensure that the user is as unconstrained as possible in their use of the technology. The main features engineered and defined are shown in Figure 6 and explained below. They are based on the acceleration components a_x, a_y, and a_z as well as the angular velocity components ω_x, ω_y, and ω_z, captured across N time samples.

Magnitude features

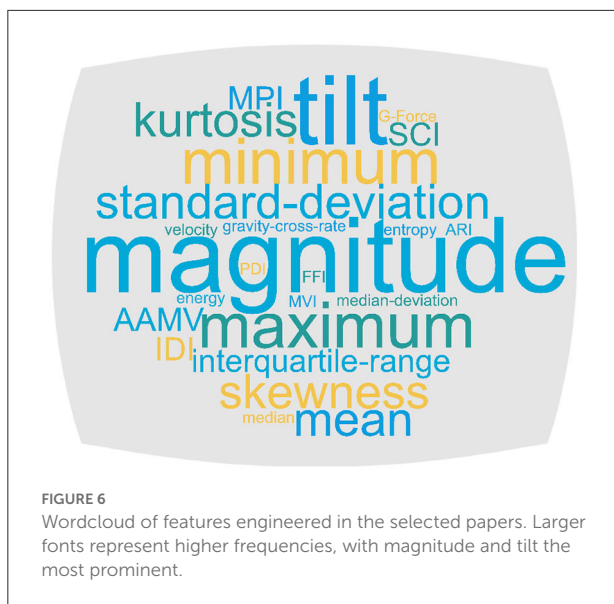Acceleration magnitude (4, 5, 8–18):
$$a_t = \sqrt{a_{t,x}^2 + a_{t,y}^2 + a_{t,z}^2}$$
G-Force (12): $G_t = \frac{a_t}{9.8}$

Differential features

Gravity cross-rate (4):
$$Z_g = \frac{1}{2N} \sum_{i=0}^{N-1} \text{sign}(a_i - g_{\text{upper bound}}) - \text{sign}(a_{i-1} - g_{\text{lower bound}})$$

Tilt (4, 8, 11): $\cos\theta_{t_1,t_2} = \frac{\vec{a_{t_1}} \cdot \vec{a_{t_2}}}{a_{t_1}\,a_{t_2}}$

Velocity (11): $\vec{v}_t = \vec{v_{t_0}} + \sum_{i=t_0}^{t} \vec{a}_i dt$

Absolute Acceleration Magnitude Variation (13, 15): $AAMV_I = \frac{1}{|I|} \sum_{i \in I} |a_{i+1} - a_i|$

Statistical features

For k an axis in x, y, z:

Maximum: $M_k = \max_{0 \le t \le N}(a_{t,k})$

Minimum: $m_k = \min_{0 \le t \le N}(a_{t,k})$

Average: $\mu_k = \frac{1}{N} \sum_{i=0}^{N-1} a_{i,k}$

Standard deviation: $\sigma_k = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (a_{i,k} - \mu_k)^2}$

Skewness: $\mu_k^{(3)} = \frac{1}{N} \sum_{i=0}^{N-1} \frac{(a_{i,k} - \mu_k)^3}{\sigma_k^3}$

Kurtosis: $\mu_k^{(4)} = \frac{1}{N} \sum_{i=0}^{N-1} \frac{(a_{i,k} - \mu_k)^4}{\sigma_k^4}$

Time domain features

These features are derived from the temporal analysis of a fall event. The following is a brief overview of a few examples; the exact definitions can be found in some of the selected articles (13, 14) and their respective references.

Impact Duration Index: IDI = duration of the fall.

Peak Duration Index: PDI = duration of the acceleration peak.

First acceleration magnitude value: FAM = time of the first acceleration peak above $2g$.

Peak time: PT = time of the peak exceeding the $3g$ threshold.

In addition to feature engineering, many of the studies (4, 5, 7, 13, 15, 16, 18) used signal filtering to smooth the signal

and remove its highest-frequency components, which represent mostly noise, prior to the detection task.

For the classification step, Figure 7 shows a categorization of approaches used between threshold-based and machine-learning-based. Six studies (8, 10–12, 16, 17) used quite basic threshold classifiers, which give a fall alert if a certain quantity (generally derived from the acceleration) exceeds a pre-defined level. While this approach is easily explainable and computationally efficient, it is likely to create many false positives in real-life scenarios. The nine other studies (4–7, 9, 12–14, 18) instead explored machine learning classifiers, with only two (15, 18) going beyond standard machine learning algorithms. The reviewed articles reportedly used the following programming languages for the implementation of their approaches: Java (6–15, 17, 18) (for Android OS), MATLAB (4, 16), and RAPIDMINER (5).

Evaluation

Commonly reported metrics for evaluating the results are sensitivity, specificity, and accuracy, which are based on the following detection outcomes:

- True Positive (TP): A fall is correctly classified as a fall
- True Negative (TN): A non-fall activity is correctly classified as non-fall activity
- False Positive (FP): A non-fall activity is incorrectly classified as a fall
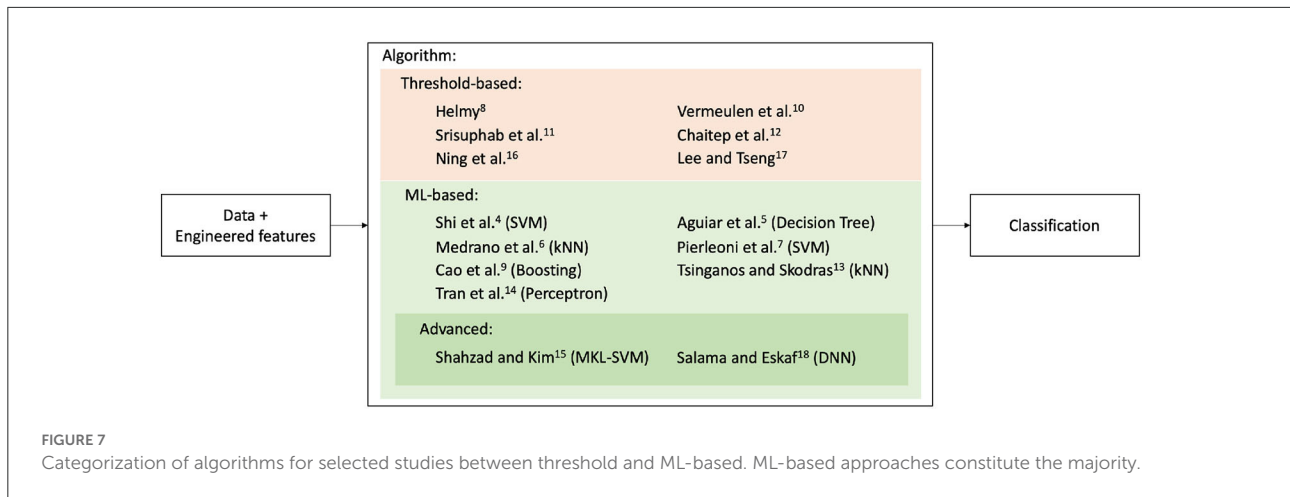- False Negative (FN): A fall is incorrectly classified as a non-fall activity

These are defined as follows for binary classification, with m the total number of samples: Sensitivity $= \frac{\text{TP}}{\text{TP+FN}}$ Specificity $= \frac{\text{TN}}{\text{TN+FP}}$ Accuracy $= \frac{\text{TP+TN}}{m}$

Sensitivity is therefore the ability to detect falls when they occur-the higher the sensitivity, the fewer falls are missed. Specificity is the ability to avoid false alarms, with a higher specificity meaning fewer false alarms-one can see a clear trade-off between sensitivity and specificity. Finally, accuracy is a mixture of both, being the ability to correctly classify an event, with no particular focus.

At first glance, the studies appear to report high metrics, with an average sensitivity of 90.6%, specificity of 91.5%, and accuracy of 92.2% and the best reaching 99.3% (7), 100.0% (17), and 99.0% (17), respectively. However, these results should be treated with caution for three reasons. First, eight of the articles only used a training set to both tune and evaluate their algorithms, greatly increasing the risk of overfitting and thus producing over-confident results. Figure 8 presents an overview of the reported sensitivities and specificities and contrasts them with the confidence in the evaluation method; it notably shows that

FIGURE 7
Categorization of algorithms for selected studies between threshold and ML-based. ML-based approaches constitute the majority.
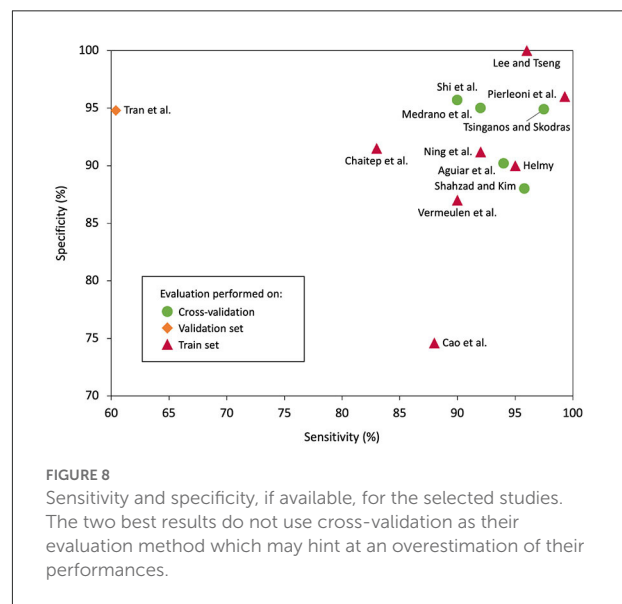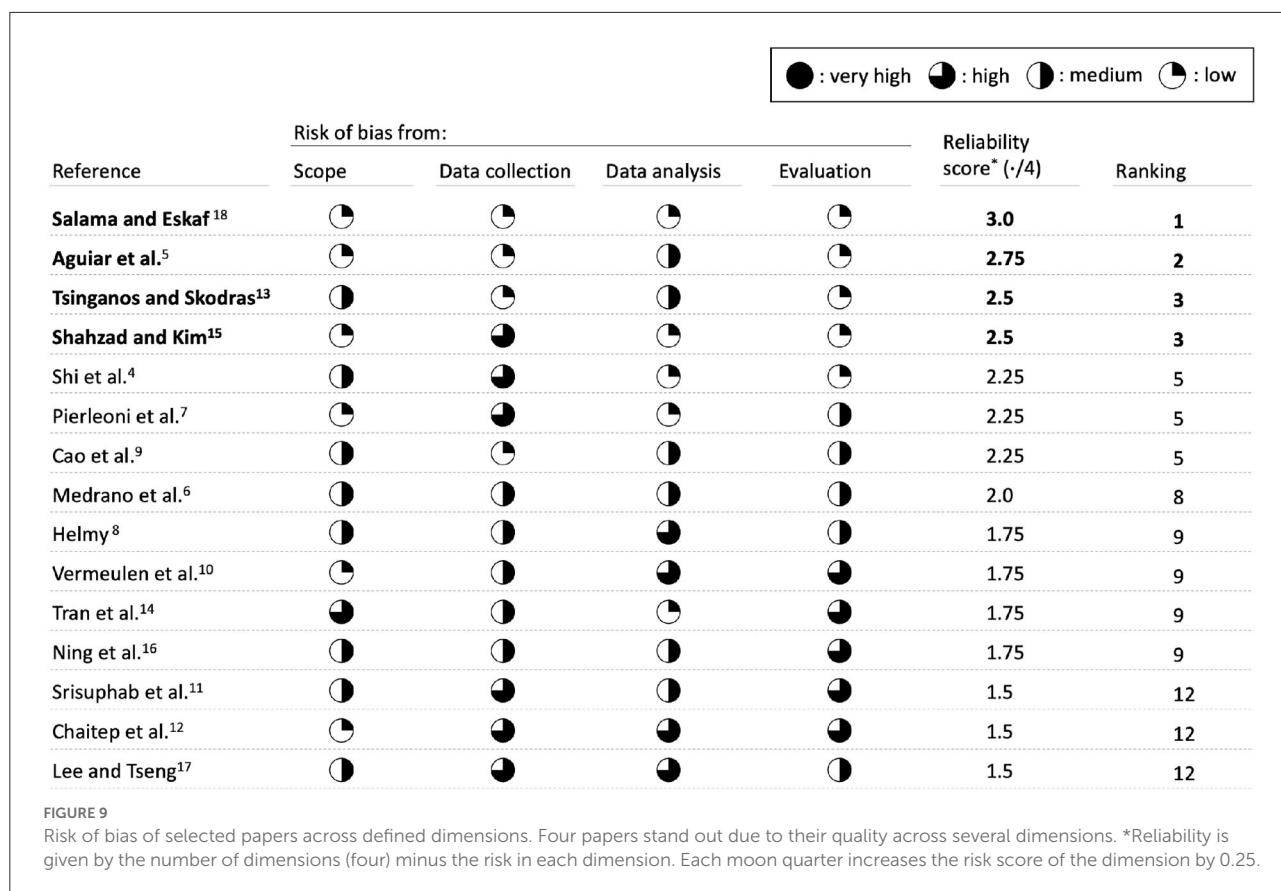
the two apparently leading results (7, 17) used an inaccurate evaluation method on the training set. Second, most of the studies used their own original datasets, which are, to different degrees, challenging to classify, depending on the scope of the daily activities considered. Third, all but one[11] of the studies evaluated their results offline-that is, the detection app was not running in someone's pocket, but instead used pre-collected data on a computer-which may have caused bias. Among the best approaches to evaluation, we find that the cross-validation by Shi et al. (4) is particularly well-considered because it is the only one that uses the leave-one-subject-out method to properly estimate how their SVM-based classification could perform with new subjects upon which it has not learned.

Finally, a thorough evaluation would consider the power consumption of the app, which only six studies reported (5, 8, 9, 13, 15, 18); it should be obvious that a smartphone battery must be able to last for a full day with the app running in the background. Power consumption can be reported using battery percentage per hour (%/h) (5, 8, 15, 18), energy per analysis (J) (9), or electric charge (mAh) (13). The range for those that reported in battery percentage per hour is 0.9–6.0%.

Bias evaluation and ranking of selected articles

Figure 9 summarizes the bias assessment of the approach of each article across the dimensions chosen for the review. Based on the criteria identified in Table 1, we estimated a comparative risk of bias for the 15 selected articles on the following dimensions: scope, data collection, data analysis, and evaluation. The risk score ranges from 0 to 1. After adding the score for each dimension, we can rank each study, as shown in Figure 9. We found Salama and Eskaf (18) to be superior according to our criteria, as they display a low risk of bias across

FIGURE 8
Sensitivity and specificity, if available, for the selected studies. The two best results do not use cross-validation as their evaluation method which may hint at an overestimation of their performances.

all dimensions. They are followed by Aguiar et al. (5), which also have three dimensions with a low bias risk but a slightly higher risk on the data analysis dimension. This is due to using a decision tree as a classification algorithm, which may lack the complexity to classify tough cases. Tsinganos and Skodras (13) and Shahzad and Kim (15) share third place. Tsinganos and Skodras (13) have a low risk on data collection and evaluation but quite a restrictive scope featuring only four types of falls. In addition, regarding data analysis, they are using kNNs which may be too simplistic in some cases. Shahzad and Kim (15) have a low bias risk on scope, data analysis, and evaluation. Conversely, they feature a high risk on the data collection, as only four participants were included in the data acquisition process, which leads eventually to low statistical significance.

	● : very high	◑ : high	◐ : medium	◔ : low

Reference	Risk of bias from: Scope	Data collection	Data analysis	Evaluation	Reliability score* (·/4)	Ranking
Salama and Eskaf [18]	◔	◔	◔	◔	**3.0**	**1**
Aguiar et al. [5]	◔	◔	◐	◔	**2.75**	**2**
Tsinganos and Skodras [13]	◐	◔	◐	◔	**2.5**	**3**
Shahzad and Kim [15]	◔	●	◔	◔	**2.5**	**3**
Shi et al. [4]	◐	●	◔	◔	2.25	5
Pierleoni et al. [7]	◔	●	◔	◐	2.25	5
Cao et al. [9]	◐	◔	◐	◐	2.25	5
Medrano et al. [6]	◐	◐	◐	◐	2.0	8
Helmy [8]	◐	◐	◑	◐	1.75	9
Vermeulen et al. [10]	◔	◐	●	●	1.75	9
Tran et al. [14]	●	◑	◔	●	1.75	9
Ning et al. [16]	◐	◐	◐	●	1.75	9
Srisuphab et al. [11]	◐	●	◐	●	1.5	12
Chaitep et al. [12]	◔	●	◑	●	1.5	12
Lee and Tseng [17]	◐	●	●	◐	1.5	12

FIGURE 9

Risk of bias of selected papers across defined dimensions. Four papers stand out due to their quality across several dimensions. *Reliability is given by the number of dimensions (four) minus the risk in each dimension. Each moon quarter increases the risk score of the dimension by 0.25.

Discussion

From our findings, we compiled four main areas of improvement for future work on fall detection apps that use smartphone accelerometers.
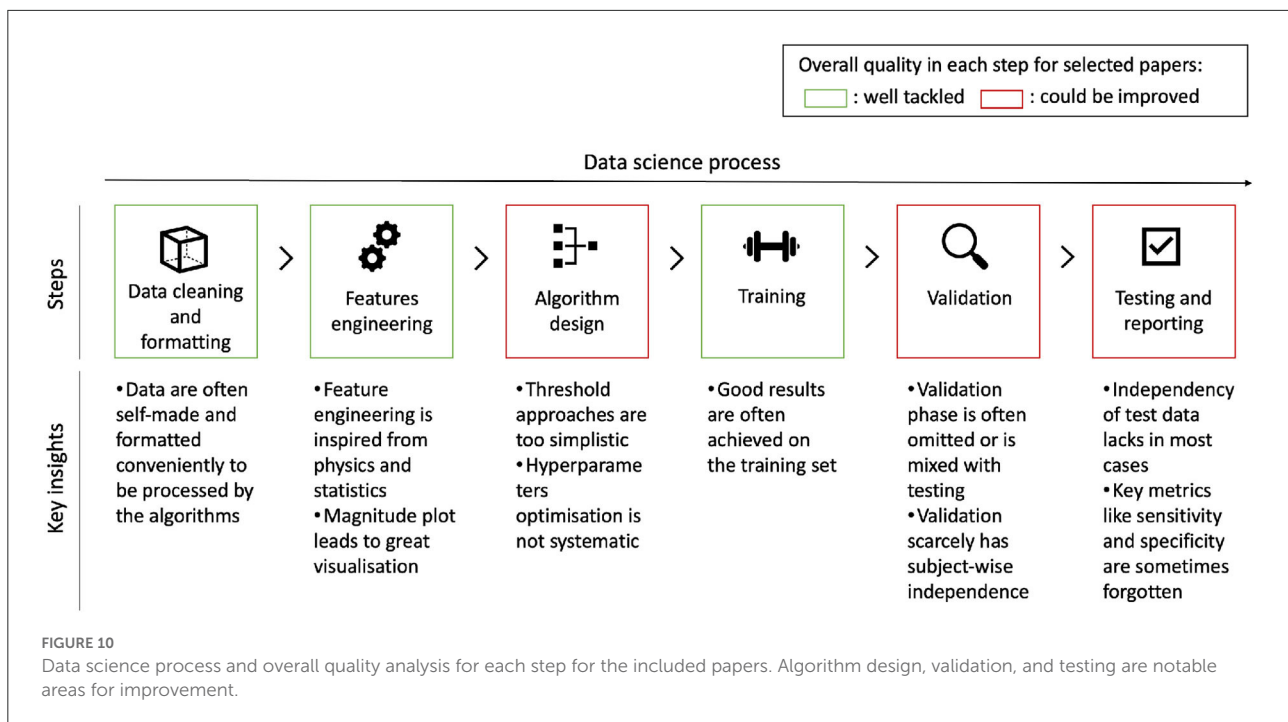
Adequation between study scope and end-users habits

Studies must account for the specifics of the population upon which they are focusing in order to replicate them in their experimental investigations. First, regarding smartphone positioning, most of the selected studies (4, 5, 9, 11–13, 15, 17) assumed that the phone would always be placed in the pants pocket. Despite being a common position, this is still a significant assumption that could be erroneous when considering the full span of a normal day. Mechanisms should be implemented to either correctly identify the real position of the device or to encourage the user to put the device where it is expected, such as by sending a notification if the phone is stationary, for instance. Second, the scope of the ADLs considered in a study should be as large-and pre-empt as many false positives-as possible. One can think of many daily actions that might cause a sudden acceleration that could be mistakenly classified as a fall, like lying down or dropping the phone. Such false positives would trigger false alarms in real life, which would deteriorate the user experience and hence the app utilization rate.

Tailored app design

The needs that the fall app will meet should be clearly identified. Only one study (4) directly asked the targeted population what key functionalities a fall application should offer. It subsequently implemented, for instance, the transmission of relevant personal information alongside the alert. Another study emphasized that, because users may not be tech-savvy, the app interface should be as curated and straightforward as possible and should feature streamlined processes (11). Besides visuals and functionality, battery usage is also critical, and to be truly usable, the smartphone battery should last at least a full day while the app is functioning. This means that battery-saving optimizations should be applied without negatively impacting the app's performance. Among the studies were multiple ideas to minimize power usage, such as leveraging an adaptive sampling frequency (5, 8, 13, 18),

FIGURE 10
Data science process and overall quality analysis for each step for the included papers. Algorithm design, validation, and testing are notable areas for improvement.

avoiding unnecessary computations (15, 18), and using fog computing (9).

Rigorousness of the data science process

Data science processes should be standardized to ensure both the relevancy and usability of study results. The first area of attention is splitting the dataset into a training set, a validation set, and a test set. A training set allows the algorithm fit to be created, the validation set is used to select hyperparameters, and the test set can then be used for final evaluation and results reporting. If several participants are involved, the data for any given participant should belong only to one of the three sets. Such an approach lays the groundwork for optimization while ensuring that algorithms have not been overfitted and are not reporting overly optimistic results. This is of particular relevance here, given the large number of varieties of fall and of people prone to falling. Figure 10 offers an overview of the overall quality in each step of the process for the selected articles. Another practice that could be systematized is the benchmarking of results. As studies often resort to building their own datasets, comparisons lose their significance. To alleviate this, studies could also report results that they achieve using publicly available datasets. For instance, the datasets *MobiAct* (26) and *UniMiB SHAR* (25) are both recent, public datasets and contain appropriate data for the scope of fall detection studies that the current paper has considered.

Advanced machine learning opportunities

Advanced machine learning methods are of interest in this research area because of their potential for delivering both good performance and robustness. Besides the two selected studies (14, 18) that employed non-basic machine learning approaches, other studies in the literature (27–29) have considered deep learning methods, and despite not creating apps and thus not being selected for this review, they still laid out interesting ideas. They chose neural architectures that are especially relevant for time series analysis-specifically, gated recurrent units (27) (GRU), long-short term memory networks (28) (LSTM), and convolutional neural networks (29) (CNN). When tested for activity recognition, GRU architecture has notably been shown to outperform LSTM, CNN, support vector machines, Adaboost, and k-nearest neighbors in terms of accuracy and in terms of F1-scores on the publicly available *MobiAct* dataset (27). The better performance of deep learning is generally explained by its ability to account for the temporality of the data compared to traditional machine learning approaches (27). Receiving feedback directly from the user through the reporting of bugs, such as false positives, has also been suggested (28) and could lead to the implementation of both online learning mechanisms and personalization. Further studies should then investigate whether such architectures could be embedded in an app, how they could be used for real-time detection, and whether they lead to excessive power consumption.

Recommendations	Reduces risk of bias on:			
	Scope	Data collection	Data analysis	Evaluation
1. Include data from targeted populations (elders, people with disabilities or recovering from an operation) or assess how well the developed detection system will generalize to them.	✔	✔		
2. Investigate the carrying habits of smartphone owners, especially bags for women and other older people, to design an adapted and inclusive data collection procedure.	✔	✔		
3. Include in the scope ADLs that feature sudden accelerations in order to separate them from falls. The algorithm should be trained with cases that are difficult to separate due to their similarity.	✔			
4. Include in the scope numerous types of falls. In real life, there is a wide range of ways of falling that should be accounted for.	✔			
5. Use deep learning architectures to fit for time series analysis. Deep learning is particularly good at extrapolation and can continue producing good results beyond the scope of the study.			✔	
6. Include contextual information to the fall detection, such as the fall peak acceleration or if a loss of consciousness occurred.		✔		✔
7. Inquire about and implement key functionalities that a fall detection app should offer. Ultimately, the value of an app is in its functionalities.				✔
8. Ensure low battery consumption through adaptive computing and sampling. The app should be able to work for at least a full day.				✔
9. Use leave-one-subject-out cross-validation to optimize the hyperparameters. This is a direct way to run an unbiased performance evaluation and to optimize.				✔
10. Evaluate and report using an independent test set in order to anticipate the behavior of the detection algorithm on unseen data.				✔
11. Benchmark results using a publicly available dataset. Comparisons using the same standards as other studies create clarity and can lead to fruitful discussions.				✔
12. Run live detection tests to identify new and unexpected technical issues, providing the opportunity to fix them.				✔

FIGURE 11
Twelve recommendations to reduce risk of bias on the scope, data collection, data analysis, and evaluation dimensions.

Recommendations for future work

We give in Figure 11 a list of twelve concise recommendations that are considered both actionable and impactful, so that even better fall detection apps with less risk of bias can be built as part of future research.

Conclusion

In the current review, we first defined relevant criteria for selecting articles closest to having a machine learning-based solution for fall detection. We identified 15 studies with an operating application and defined categories to structure our analysis, notably highlighting narrow scope, uneven depths of data analysis, and possible biases in the evaluation process. We finally constructed 12 actionable and promising recommendations for future studies, which should give more consideration to end-user specifics, data science methodology, and more advanced machine learning.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

ME designed and led the study. TS, ME, RF, and CM conceived the study. All authors approved the final manuscript.

Funding

Conflict of interest

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

Publisher's note

References

1. Falls. World Health Organization. In: *WHO Fact sheets [website]* (2021). Available online at: https://www.who.int/news-room/fact-sheets/detail/falls (Accessed April 06, 2022).

2. World Health Organization. *Step Safely: Strategies for Preventing and Managing Falls Across the Life-Course*. World Health Organization (2021).

3. Ageing and health. World Health Organization. In: *WHO Fact Sheets [website]* (2021). Available online at: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health (Accessed April 06, 2022).

4. Shi Y, Shi Y, Wang X. Fall detection on mobile phones using features from a five-phase model. In: *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*. Fukuoka, Japan (2012). p. 951–956.

5. Aguiar B, Rocha T, Silva J, Sousa I. Accelerometer-based fall detection for smartphones. In: *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. Lisboa, Portugal (2014). p. 1–6.

6. Medrano C, Igual R, Plaza I, Castro M, Fardoun HM. Personalizable smartphone application for detecting falls. In: *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. Valencia, Spain (2014). p. 169–72.

7. Pierleoni P, et al. SVM-based fall detection method for elderly people using android low-cost smartphones. In: *2015 IEEE Sensors Applications Symposium (SAS)*. Zadar, Croatia (2015). p. 1–5.

8. Helmy A, Helmy A. Seizario: novel mobile algorithms for seizure and fall detection. In: *2015 IEEE Globecom Workshops (GC Wkshps)*. San Diego, CA (2015). p. 1–6.

9. Cao Y, Chen S, Hou P, Brown D. Fast: a fog computing assisted distributed analytics system to monitor fall for stroke mitigation. In: *2015 IEEE International Conference on Networking, Architecture and Storage (NAS)*. Boston, MA (2015). p. 2–11.

10. Vermeulen J, Willard S, Aguiar B, De Witte LP. Validity of a smartphone-based fall detection application on different phones worn on a belt or in a trouser pocket. *Assist Technol.* (2015) 27:18–23. doi: 10.1080/10400435.2014.949015

11. Srisuphab A, Silapachote P, Phongpawarit J, Visetpalitpol S, Jirapasitchai S. Redle: Elderly care on clouds. In: *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. Khon Kaen, Thailand (2016). p. 1–4.

12. Chaitep T, Chawachat J. A 3-phase threshold algorithm for smartphone-based fall detection. In: *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. Phuket, Thailand (2017). p. 183–6.

13. Tsinganos P, Skodras, A. A smartphone-based fall detection system for the elderly. In: *Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis*. Ljubljana, Slovenia (2017). p. 53–8.

14. Tran HA, Ngo QT, Tong V. A new fall detection system on android smartphone: Application to a sdn-based iot system. In: *2017 9th International Conference on Knowledge and Systems Engineering (KSE)*. Hue, Vietnam (2017). p. 1–6.

15. Shahzad A, Kim K. Falldroid: an automated smart-phone-based fall detection system using multiple kernel learning. *IEEE Trans Indus Inform.* (2018) 15:35–44. doi: 10.1109/TII.2018.2839749

16. Ning Y, Hu S, Nie X, Liang S, Li H, Zhao G, et al. Real-time action recognition and fall detection based on smartphone. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Honolulu, HI (2018). p. 4418–22.

17. Lee JS, Tseng HH. Development of an enhanced threshold-based fall detection system using smartphones with built-in accelerometers. *IEEE Sensors J.* (2019) 19:8293–302. doi: 10.1109/JSEN.2019.2918690

18. Salama MS, Eskaf K. An intelligent mobile app for fall detection. In: *2020 30th International Conference on Computer Theory and Applications (ICCTA)*. Virtual conference (2020). p. 65–70.

19. Insel TR. Digital phenotyping: technology for a new science of behavior. *JAMA.* (2017) 318:1215–6. doi: 10.1001/jama.2017.11295

20. Van Laerhoven K, Borazio M, Burdinski JH. Wear is your mobile? investigating phone carrying and use habits with a wearable device. *Front ICT.* (2015) 2:10. doi: 10.3389/fict.2015.00010

21. Zeleke BM, et al. Mobile phone carrying locations and risk perception of men: a cross-sectional study. *PLoS ONE.* (2022) 17:e0269457. doi: 10.1371/journal.pone.0269457

22. Redmayne M. Where's your phone? a survey of where women aged 15-40 carry their smartphone and related risk perception: a survey and pilot study. *PLoS ONE.* (2017) 12:e0167996. doi: 10.1371/journal.pone.0167996

23. Casilari E, Luque R, Morón MJ. Analysis of android device-based solutions for fall detection. *Sensors.* (2015) 15:17827–94. doi: 10.3390/s150817827

24. Wang Z, Ramamoorthy V, Gal U, Guez A. Possible life saver: a review on human fall detection technology. *Robotics.* (2020) 9:55. doi: 10.3390/robotics9030055

25. Micucci D, Mobilio M, Napoletano P. Unimib shar: a dataset for human activity recognition using acceleration data from smartphones. *Appl Sci.* (2017) 7:1101. doi: 10.3390/app7101101

26. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M. The mobiact dataset: recognition of activities of daily living using smartphones. In: *International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 2. Prague, Czech Republic: SciTePress (2016). p. 143–51.

27. Wu X, Zheng Y, Chu CH, Cheng L, Kim J. Applying deep learning technology for automatic fall detection using mobile sensors. *Biomed Signal Process Control.* (2022) 72:103355. doi: 10.1016/j.bspc.2021.103355

28. Astriani MS, Heryadi Y, Kusuma GP, Abdurachman E. Long short-term memory for human fall detection based gamification on unconstraint smartphone position. In: *2019 International Congress on Applied Information Technology (AIT)*. Yogyakarta, Indonesia (2019). p. 1–6.

29. Deep S, Tian Y, Lu J, Zhou Y, Zheng X. Leveraging multi-view learning for human anomaly detection in industrial internet of things. In: *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. Virtual conference (2020). p. 533–7.