sensors [51,52,15,20,27,16], in which algorithms detect the occurrence of falls using the inclination of the sensor, since falls often result in an inversion of the polarity of the acceleration vector in the trajectory direction [53]. Another widely studied method is the use of monitoring cameras to detect falls [54,55], in which computer vision methods are employed to detect falls from human-pose estimation. In this review we investigate only wearable sensors, so vision-based approaches are outside our scope.

Regarding assessment of fall risk, there is still no consensus on a methodology, since this can be achieved either by using the recent history of falls, or by estimating the fall risk independently of the person's recent history of falls. The first approach, faller classification, is related to retrospective falls, that is, to identify individuals that experienced a fall event in the past [18,19,21,56]. This is a proxy measure for fall risk since it is known that a faller has higher risk of falling again. The second strategy, fall risk screening, should be prospective, that is, considering only the risk of future falls [27,5,57]. In this case, the person may not have suffered a fall in the recent history and the method should still be able to assess fall risk.

In order to assess fall risk, some studies use scores obtained from functional tests to estimate risk of falls, without the need for conducting a prospective search [58,37,25,30]. The signal acquisition may be performed while volunteers perform functional tests, predefined tasks of daily living, or in a continuous and unconstrained manner. There is a variety of feature extraction techniques and methods to produce the fall risk estimation in the literature. While some papers employ machine learning to distinguish fallers and non-fallers [38,15,17,21], others employ statistical analyses to compare features between groups of fallers and non-fallers, which may be formed retrospectively or prospectively [5,25,28,29]. Therefore, considering the heterogeneity of the studies regarding the methodology of acquisition, processing, application and signal analysis, there is a lack of consensus about the parameters that are best suited to study falls, especially in older persons [59]. Besides, according to Van Schooten et al. [60], for a given individual, there can be meaningful differences when the analysis is performed in a retrospective or a prospective way.

Considering that there are no reviews addressing the three aforementioned applications regarding falls in elderly, this paper intends to fulfill this gap, by identifying the state of art of fall detection and risk assessment in older people using wearable sensors, as well as the main characteristics of the studies in the literature, benefiting future studies. In particular, this systematic review investigates: the type of sensors used and their sampling rate, the type of signal and data processing employed, the scales and tests used in addition to the signals or while the data is acquired, and the type of application/task under study.

2. Method

A systematic review design was used based on the PRISMA recommendation. The guiding steps for this review are described below:

2.1. Search strategy

Potentially relevant articles on the study of fall-related events (i.e. fall detection, faller classification or fall risk assessment or prediction) in elderly population through inertial sensors were identified via a literature search in PubMed, IEEEXplore and Scopus.

The following search string was employed:

(”accelerometer” OR’ ’inertial sensors” OR’ ’gyroscope” OR ”magnetometers”) AND (”fall” OR’ ’fall detection” OR’ ’fall risk”) AND (”gait analysis” OR’ ’signal processing” OR’ ’feature extraction”)

This search string was validated by observing the retrieval of already known to be important papers. Additional papers, considered relevant to this review, and not appearing in 75 the search were manually included.

2.2. Selection criteria

Articles were **included** in this review if they met all of the following criteria:

1 Open access article;
2 Full original papers published in peer-reviewed journals between August 2002 and June 2019 in English, Portuguese, Spanish or French languages (those the authors are qualified to understand a scientific text);
3 Articles that presented the keywords defined by the search string on the abstract or title;
4 Articles that presented methodology related to studying fall events, such as fall detection, faller classification or fall risk assessment or prediction;
5 Articles that used inertial sensors combined with functional tests and/or daily life activity to extract features for the investigation of falls;
6 Sample composed only by elderly people (60 years or more).

2.3. Exclusion criteria

1 Papers without participants age specification;
2 Review papers.

Two authors independently assessed the suitability and methodological quality of the papers. We did not include in this review articles related to the recognition of health conditions through gait, only articles related to falls events.

2.4. Study selection

The selection process is described in Fig. 1. First, in the identification step, articles were identified via database search and manual inclusion. Second, in screening step, duplicates found in different databases were removed. In the eligibility step, articles were removed following the inclusion and exclusion criteria. Finally, the included step indicate the remaining articles that are analyzed in this review.

2.5. Data analysis and quality analysis of articles

To facilitate the data analysis, the following questions were considered:

- Sensors: we were interested in study which sensor types were employed, what were the sampling rates used to acquire the signals and the position of the sensors in the user's body during signal acquisition;
- Signal processing, analysis and recognition: we investigated the most common types of data (pre)processing and method for analysis and recognition used in articles in order to achieve the outcome. This depends on the task the study tries to solve, which may be a statistical analysis of the signal characteristics, a system for classification, and others. We classified those into two main categories of signal processing: feature extraction and machine learning;
- Sample characteristic: the characteristics of the sample studied by the selected articles are an important matter due to our interest in elder populations. In particular, we were interested in studying whether the samples consisted of healthy elderly or frail elderly;
- Tests: since there are different setups for signal acquisition, it is fundamental to investigate which tests (functional or based on activities of daily living) were used in combination with the use of sensors to extract gait characteristics of the volunteers;
- Applications: finally, the type of task the papers studied is also of great interest. We classified those into three main categories of applications — fall detection, faller classification or fall risk assessment or prediction.

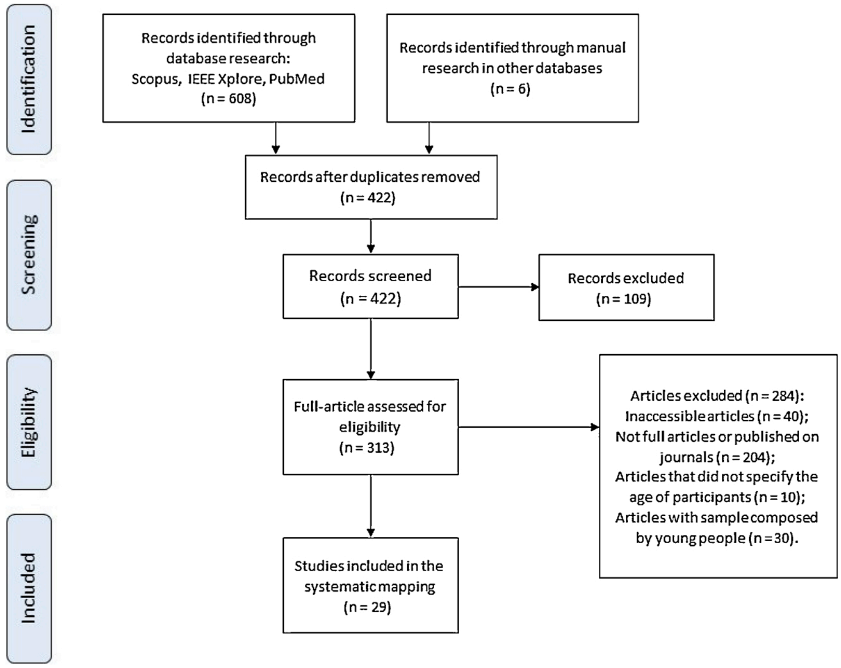

Fig. 1. Flowchart in the PRISMA standard used in the selection of article for systematic review.

2.6. Bias control

Three databases (PubMed, IEEEXplore and Scopus) were used to minimize publication bias. Moreover, just one type of publication was searched: scientific papers.

3. Results

The search strategy retrieved 608 papers, and 6 papers were manually included after searching other databases. After the exclusion of duplicated articles found in different sources, 422 remained in review. From those, 109 were removed after screening the abstract against the eligibility criteria, which left 313 papers to be read in full-text. 284 were excluded after full-text reading due also to inclusion/exclusion criteria. Therefore, 29 papers were shortlisted for this review [61,5,10,23,15,24–26,16,27,17,28,18,19,11,29,12,30–36,20,37,38,21,13]. The PRISMA flowchart, including the steps considered for the selection of articles in this review is shown in Fig. 1.

Publication year: in Fig. 2 we plot the time distribution of the articles analyzed in this review. We note that there is no defined pattern or trend over the distribution of their throughout the years. However, we observed a larger number of papers published in the years 2017–2018 when compared to previous ones [61,5,10,26,16,27,17,28,18,34,37,38,21].

Summary of papers selected for the review: Table 1 summarizes the articles selected for analysis in this review, with respect to the author and year of publication, type of application on falls, type of sensor (with employed sampling rate in Hz) and tests used for collecting data.

In the following sections we discuss each one of the questions for data analysis in more depth.

3.1. Sensors

Four type of sensors were identified in the studies as ways of mapping movement produced by the change of velocity or body patterns in the extracted signals: accelerometer, gyroscope, magnetometer and barometer. It is important to note that some articles used the combination of two or more sensors to capture motion signals [61,10,23,28,19,34,36]. The accelerometer was the choice in 28 of the 29 articles (See Fig. 3). Considering pairs of sensors, accelerometer and

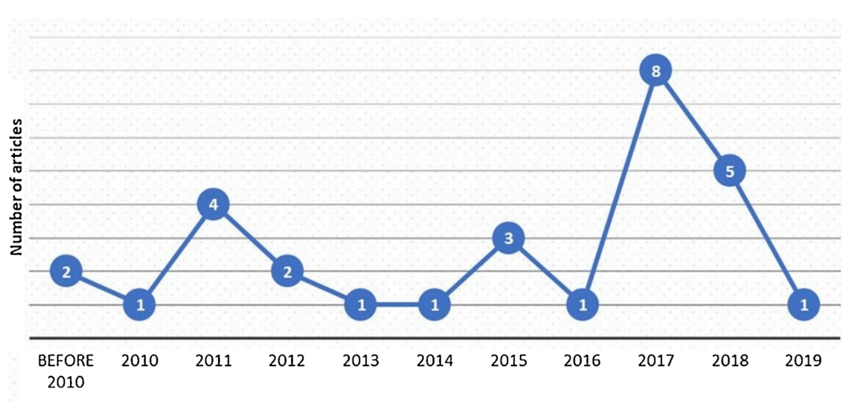

Fig. 2. Number of studies per year of publication.

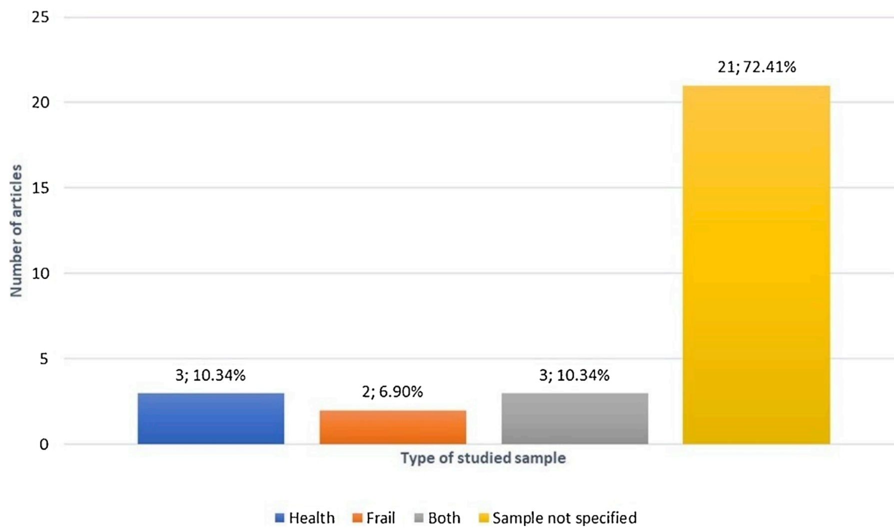

Fig. 7. Health condition of the sample used per article.

3.4. Studied sample

Another topic studied was the type of study sample in terms of participants health status. Bio-mechanical characteristics and gait traits may change significantly when considering healthy (often community-dwelling) elderly in comparison with frail (often hospitalized or institutionalized), therefore requiring specific research efforts. The 29 articles selected had an average of $n = 66$ participants ranging from $n = 3$ to 296, and only 8 reported the characteristics of the sample in depth. Fig. 7 shows the distribution of studies with this respect: three performed exclusively with healthy elderly [10,18,19], two only with frail elderly [15,11], three with both (healthy and frail elderly) [29,35,37] and 21 did not specify their sample [9,5,23–26,16,27,17,28,12,30–34,36,20,38,21,13].

3.5. Tests and scales

In relation to the tests performed with the sensors to extract the patterns, specific functional tests were used in 17 of the 29 articles [23,15,24,26,16,27,17,18,29–33,35,20,37,21]. The tests used were often those related to fall risk screening and balance, such as: Timed Up and Go test, Tinetti score, Sit-to-stand, and others. All functional tests and the frequency those are employed in the reviewed papers are described on Table 2.

On the other hand, the activities of daily living, such as going up and down stairs, and also monitoring the daily routine for consecutive days through the sensor were used in 9 studies [9,5,10,25,19,11,12,38,13] (See Fig. 8). Note that some articles used both strategies [28,34,36]. Of the 12 articles that acquired their signals through activities of daily living, 4 were made with controlled activities, that is, through the execution of daily tasks specified by the evaluator, sit and stand up [5,25,34,13], and 8 were made from signals obtained during the daily routine of the volunteer, without any kind of supervision [9,10,28,19,11,12,36,38] (Fig. 9).

3.6. Application: fall detection, faller classification or fall risk screening

There are three types of detection in relation to falls: identifying a fall that has already occurred, identifying fallers and identifying the risk of falls. In this section we try to find out which type of application was

Table 2
Full list of used functional tests with a short description and frequency of use in the reviewed papers.

Functional Test	Description	Frequency
Timed up and Go test [65]	Times the seconds taken by the volunteer to get up without backing from a chair; walk three meters straight; to turn; walking back the three meters; and sit in the chair also without support.	10
Tinetti testor Performance-oriented mobility assessment [66]	Assesses fall risk through the assessment ofbalance and performance-oriented gait.	5
Sit-to-stand [67]	Consists of performing sit-to-stand transfers as fast as possible with the arms crossed.	4
25 minute walkingtest [15]	Consists of timing the timeit takes the individual to walk 25 m	4
Alternate-step test(ATS) [68]	Placing, alternatively and rapidly,the whole of each foot onto and off of a platform (19 cm high and 40 cm wide).	3
Six-minute walk test(6MWT) [69]	Measures the distance after walking during sixminutes on a hard and flat surface at a comfortable speed without speaking.	2
Mini Motor Test(MMT) [70]	A 20-item score composed by:abilities in bed, quality of sitting position, abilities in the standing position and quality of gait.	1
Physical PerformanceScale (PPS) [71]	Evaluates the functions of the lower extremitiesby means of gait, balance and stand up tasks.	1
Fukuda Test [72]	Used for vestibular impairments detection, consists of stepping in place with outstretched arms and blinded eyes	1
One leg balance [73]	Evaluates the risk of falls and consists of staying staying in one leg for 5 seconds without help.	1
Romberg test [74]	Evaluates the static balance. The individual needs to stand for 30 seconds with eyes open on the balance platform and then repeat it with the eyes closed.	1
Berg Balance Scale [75]	Assesses functional balance performance based on 14 common items for daily living.	1
Short physical performancebattery (SPPB) [76]	Evaluates fall risk by measuringgait, balance and muscular strength	1

Table 3
Quantitative summary of the investigated questions on the reviewed papers grouped by application.

Number of papers		Fall detection n = 5	Faller classification n = 7	Fall risk screening n = 17
Sensors	Accelerometer	5	7	16
	Gyroscope	2	0	4
	Magnetometer	1	0	0
	Barometer	0	1	1
Sampling Rate (Hz)	Minimum	8	30	4
	Maximum	200	200	256
	Average	102	69	80
Sensors place	Waist	2	1	0
	Lower back	2	0	5
	Pelvis	0	3	1
	Head	0	2	1
	Ankle	0	2	2
	Others	1	6	6
Signal processing	Feature extraction	4	4	10
	Machine learning	1	3	7
Sample (n)	Minimum	3	18	11
	Maximum	38	100	296
	Average	23	56	83
Sample type	Health	1	2	3
	Frail	1	1	3
Evaluation	Functional tests	0	6	14
	ADL	5	1	6

value to a correct prediction, regardless of its category (positive or negative):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN};$$

- Sensitivity (or true positive rate): the number of true positive outcomes (fall, faller or future fall correctly predicted), divided by the sum of all positive instances:

$$\text{Sensitivity} = \text{TPR} = \frac{TP}{TP + FN};$$

- Specificity (or true negative rate): the number of true negative outcomes (not fall, non-faller or lack of future fall correctly predicted), divided by the sum of all negative instances:

$$\text{Specificity} = \text{TNR} = \frac{TN}{TN + FP};$$

- Receiver operator characteristic curve (ROC): one may set different threshold values over a variable or tool. This cut-off is often use to enforce the system to favour sensitivity over specificity and vice-versa. The ROC is a curve that represents the relationship between the sensitivity (TPR) and specificity (TNR) of a given variable or diagnostic test over a series of cut-off values. It is a relevant analysis once it is often preferable to design systems that favour Sensitivity over Specificity, so it would allow to study the optimal operating point for a given problem;
- Area under the ROC curve (AUC): considering all evaluated cut-off values, it represents the probability that the test or system will rank a randomly chosen positive instance higher than a randomly chosen negative instance, therefore correctly classifying them. Note that, while the Accuracy, Sensitivity and Specificity are computed for a single cut-off values, the AUC is a measure considering a series of increasing threshold points and that is why it is often used to evaluate diagnostic systems.

For a more detailed explanation of such metrics and their relationships please refer to Fawcett et al. [77].

A total of 17 articles performed this type of evaluation, including analysis of accuracy, sensitivity, specificity and area under the ROC curve [61,15,27,17,28,18,11,29,12,30,33–36,38,21,13]. Four articles related to fall detection reported these measures, and amongst them two articles mentioned the AUC: one of 0.796 [9] and other of 0.918 [11]; and two indicated the sensitivity and specificity. Gietzelt et al. [12] identified sensitivity of 0.913 and specificity of 0.95 while Saadeh et al. [13] found sensitivity of 0.986 and specificity of 0.993. Regarding faller classification, four articles measured the accuracy of their features: the sensitivity of the methods varied of 0.605 to 1, the specificity ranged from 0.80 to 0.84 [15,17,18,21], only Ponti et al. [18] observed the AUC (AUC = 0.84) and the accuracy 278 found by Drover et al. [17] and [21] was 0.73 and 0.79, respectively.

For the nine articles about fall risk screening [27–30,33–36,38], the accuracy ranged from 0.57 to 0.90 [78,33]. The best sensitivity result was 0.931 [34], specificity was 1 [33] and AUC was 0.75 [38].

4. Discussion

Considering the strategies for fall investigation, the accelerometer placed at the waist and/or lumbar, with a sampling rate of 50 and 100 Hz, was the most used sensor setting to capture gait signals. Regarding signal processing, the methods that directly compare features extracted from the accelerometry signal are still the most used, but an increase of machine learning methods was observed.

Besides the accelerometer, gyroscope, magnetometer and barometer data were also used in the studies. Together, accelerometer and gyroscope were employed in 17.2% of articles, and accelerometer and barometer were observed at 6.9%. This fact corroborates with Howcroft et al. [79], that found the most used sensors for fall risk assessment research are accelerometer and gyroscope, with 27.5% of the articles combining these two sensors. However, this considers studies for all population, not only on older person.

The results show that the waist and lumbar region were the most used. According to Ozdemir [52] the waist region is considered the most suitable location providing 99.96% sensitivity for detection of falls. Howcroft et al. [79] have identified that regions near the center of mass, such as the lumbar and the waist, are superior for identifying risk of falls when compared with other sites. Montesinos et al. [80] also identified the lumbar region as the most common positioning of the sensors, however the authors did not identify the waist region as one of them. Already Schwickert et al. [9] observed that the waist or hip region were the most used for detection of falls.

A sampling rate of 50 Hz or 100 Hz were the most observed, but we found this parameter ranging from 4 to 256 Hz. This result confirms the one found by Schwickert et al. [9], who identified that most articles used 100 Hz or more for fall detection sampling. Pogorel and Gams [81] observed in their study on the use of wearable inertial sensors in biomechanical studies that the accelerometer, gyroscope and magnetometer sampling rates range from 20 to 800 Hz. One study [33] employed a sampling rate that would fail in achieving the minimum frequency required to measure human movement, which is 30 Hz according to Antonsson and Mann [82]. Other study [21] used exactly 30 Hz of signal sampling and even with 30 Hz, the guaranteed 15 Hz component for analysis would have only two data points per cycle, and would be a rough estimate [82]. The wide use of 50 to 100 Hz sampling rate may be explained by the fact that those are default sampling rates often found in sensors that are commercially available for acquiring this type of data [83]. Lower values are often associated with systems that are implemented using general-purpose wearable or mobile devices, such as smartwatches and smartphones, that seldom have capacity for sampling over 50 Hz.

The feature extraction was found in this review as the most common form of type of signal processing, being the features comparison