Review

Wearables and Machine Learning Use for Activity of Daily Life and Fall Management in Older Adults: Systematic Review

Izabela Pera Calvi[1], MD; Filipe Andrade Bernardi[2], PhD; Francisco Ferreira[1], MSc; Vinícius Costa Lima[2], PhD; Isabela Carvalho Guerche[1], MSc; Tiago Silva[1], MSc; Duarte Azevedo Moura Dias[1], MSc; Teresa Valverde da Silva[1], MSc; Maria Cunha Ribeiro Rocha[1], MSc; Rute Almeida[2], PhD; Manuel Fernández-Delgado[3], PhD; Alberto Freitas[2], PhD

[1]MEDCIDS, Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal

[2]RISE-Health, Faculty of Medicine, University of Porto, Porto, Portugal

[3]Centro Singular de Investigación en Tecnoloxías Intelixentes da USC (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain

Corresponding Author:

Alberto Freitas, PhD
RISE-Health, Faculty of Medicine
University of Porto
Alameda Prof. Hernâni Monteiro
Porto 4200-319
Portugal
Phone: +351 220 426 534
Email: alberto@med.up.pt

Abstract

Background: Population aging has increased interest in digital strategies that may support functional independence and safety in older adults. Consumer-grade wearable technologies combined with machine learning have been increasingly studied for monitoring activities of daily living, mobility assessment, fall-related applications, and other autonomy-relevant functions.

Objective: This systematic review aimed to identify, appraise, and synthesize primary studies on consumer-grade wearable technologies combined with machine learning in older adults, focusing on application domains, methodological characteristics, validation strategies, and reported outcomes, and relevance to autonomy, daily functioning, mobility monitoring, and fall-related safety.

Methods: We conducted a systematic review in accordance with the Cochrane handbook and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to synthesize evidence on the use of wearable technologies combined with AI, particularly machine learning methods, to support daily living and fall-related safety among older adults. The search was revised and conducted in PubMed/MEDLINE, Scopus, Institute of Electrical and Electronics Engineers Xplore, and the Association for Computing Machinery Digital Library from database inception through May 2026, with no date or language restrictions. Eligible reports were full primary studies, including conference proceedings when sufficient methodological detail and results were available for assessment. Search reporting was aligned with PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension).

Results: Thirty studies were included. Most studies were observational or methodological and relied primarily on inertial sensing, with smartphone and mobile-sensing approaches contributing to specific subdomains. The main application domains were activities of daily living monitoring, gait and mobility assessment, cognitive/dementia-related monitoring, Parkinson-related monitoring, fall detection or fall-risk assessment, and frailty assessment. Classical machine learning models, such as support vector machines and random forests, and deep learning architectures, such as convolutional neural networks and long short-term memory networks, were commonly used. However, the evidence base was highly heterogeneous, frequently based on small samples, and rarely supported by external validation or clinically meaningful outcomes.

Conclusions: Consumer-grade wearable technologies combined with machine learning appear technically promising for monitoring autonomy-relevant domains in older adults, including daily functioning, mobility, and fall-related applications. However, the current evidence remains exploratory and is limited by substantial methodological heterogeneity, small sample sizes, limited external validation, and infrequent reporting of clinically meaningful or patient-centered outcomes. Currently, the literature supports technical feasibility more clearly than robust real-world generalizability or routine clinical adoption. Future research should prioritize free-living evaluation, transparent reporting, external validation, interdisciplinary co-design, and patient-centered outcomes to strengthen real-world relevance.

Trial Registration: PROSPERO CRD420251044449; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251044449

JMIR AI 2026;5:e93640; doi: 10.2196/93640

Keywords: older adults; autonomy in daily living; wearable technology; machine learning; artificial intelligence

Introduction

Recent projections for the European Union indicate that by 2050, older adults will account for nearly one-third of the population, with fewer than 2 working-age individuals per older adult [1]. This increase is a milestone of our century, driven by technological and medical advances in health care promotion, including sanitary measures, higher vaccination rates, and the development of new antibiotics [2]. It demonstrates a reduction in the duration, severity, and incidence of major diseases [3]. However, the substantial increase in the older population is already affecting countries' productivity and placing additional pressure on public finances, given the rising pension obligations, greater reliance on health services, and the need for continuous intensive care. This dynamic has become a significant economic and public health concern [2].

Therefore, with increasing life expectancy, health issues such as frailty, dementia, cardio- and cerebrovascular diseases, and mobility limitations become more prevalent [4]. Atrial fibrillation is increasingly common and a major contributor to cerebrovascular events, thereby increasing the need for continuous support in this population throughout the day [3]. Thus, these older adults should be able to maintain autonomy in safely performing their daily tasks, reducing the burden on health care services and decreasing dependence on caregivers.

One promising approach is to integrate emerging technologies, such as wearable devices, machine learning (ML), and smart sensors, to support older adults in their daily activities. Wearable devices encompass a broad range of technologies, including fitness trackers, smartwatches, smart rings, smart glasses, smart clothing, and medical wearables. In the current review, however, the focus was specifically on consumer-grade wearable technologies (CGWT) and related mobile sensing platforms that are broadly available outside specialized medical-device contexts. Moreover, each wearable device can contain one or multiple sensors. The sensors are diverse, can detect different vital signs, and allow the estimation of physiological parameters such as heart rate (HR), heart rhythm, blood pressure (BP), oxygen saturation (SpO_2), skin temperature, physical activity levels, sleep patterns, and others [5].

These technologies can promote independence by enabling the safe execution of daily activities and allowing continuous monitoring. By using ML and AI to analyze the collected data, they can detect early health risks and support timely responses that help older adults remain autonomous and safe [5].

Previous reviews have examined adjacent topics, including AI-enabled wearable sensing for digital health, disease-specific digital biomarkers, and broader applications of ML in neurological and aging-related contexts [6]. Nevertheless, these reviews have not provided a focused, systematic synthesis of studies specifically addressing CGWT combined with ML in older adults across domains directly related to autonomy, such as activities of daily living, mobility, fall-related safety, and functional monitoring [5]. Moreover, the available review literature has generally been limited by either a broader technological focus, a narrower disease focus, or insufficient attention to validation strategies and to the distinction between technical performance and clinically meaningful utility in older populations [7]. Accordingly, this systematic review (SR) aimed to systematically identify, critically appraise, and synthesize primary studies examining CGWT combined with ML in older adults. The review focused on application domains, methodological characteristics, validation strategies, and reported outcomes, with particular attention to the relevance of these technologies for autonomy, daily functioning, mobility monitoring, and fall-related safety.

Methods

Review Design and Reporting Framework

This SR was conducted in accordance with the Cochrane Handbook for Systematic Reviews of Interventions and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [8,9]. The review synthesized evidence on consumer-grade technologies combined with ML with a focus on application domains, methodological characteristics, validation strategies, and reported outcomes, and relevance to daily living, mobility monitoring, and fall-related safety. Particular attention was given to methodological gaps, real-world applicability, and the distinction between technical model performance and clinically or functionally meaningful outcomes.

The review protocol was prospectively registered in PROSPERO (International Prospective Register of Systematic Reviews) under the registration number CRD420251044449. By documenting the predefined objectives, eligibility criteria, and analytical strategies, this prospective specification enhances methodological transparency. It constrains post hoc methodological changes and selective reporting, thereby minimizing reporting bias throughout the review process. No deviations from the registered protocol occurred. No amendments were made to the protocol after registration [8].

This review was conducted and is reported as an SR because it addressed a prespecified review question, applied predefined eligibility criteria, used a systematic multidatabase search, performed study selection and data extraction in duplicate, assessed methodological quality using a structured risk-of-bias tool, and synthesized findings across included primary studies [9]. Given the expected methodological diversity, the review was planned as an SR with structured narrative synthesis. The rationale for not conducting meta-analysis is detailed below.

Review Question

The review question was prespecified to guide the synthesis of primary studies on the combination of CGWT and ML in older adults. Given the expected heterogeneity of study designs, we used a structured eligibility framework centered on population, index technology, and outcomes of interest to ensure focus and reproducibility [10].

The framework was defined as follows: population: older adults aged 60 years or older, regardless of health status or living arrangement; index technology or phenomenon of interest: CGWT (eg, smartwatches, smart bands, smart rings, and smartphones and comparable nonspecialized mobile or wearable sensing platforms) used in combination with ML or AI methods for monitoring, classification, prediction, or decision support [5]; comparator or reference standard: when available, any reference standard, alternative assessment, clinician judgment, annotation procedure, or comparator method reported by the primary study; and outcomes or domains of interest: outcomes or study targets related to activities of daily living, mobility, fall detection or fall risk, frailty, cognitive- or Parkinson-related functional monitoring, functional independence, quality of life, usability, acceptability, validation performance, or clinical/service relevance [4,10]. These outcomes and domains were used to define eligibility and structure the narrative synthesis. They were not treated as a predefined set of intervention effectiveness end points for estimating pooled clinical effects.

For this review, CGWT were defined as wearable or mobile-connected devices that are widely available to the general public or positioned within consumer digital health markets, such as smartwatches, smart bands, smart rings, smartphones, and comparable nonspecialized devices with embedded sensing capabilities [5]. The term was used to distinguish these devices from research prototypes, custom-built experimental systems, and specialized medical-grade hardware intended primarily for restricted clinical or laboratory use [7]. Studies were considered eligible when the wearable component reflected a consumer-grade platform, even if the analytical pipeline, software environment, or data-processing framework was research oriented [11].

Accordingly, the main review question was: How have CGWT combined with ML been studied in older adults, in terms of application domains, methodological characteristics, validation strategies, and reported outcomes, and relevance to autonomy, daily functioning, mobility monitoring, and fall-related safety?

Eligibility Criteria

Studies were considered eligible if they met all of the following criteria: (1) included older adults aged 60 years or older, regardless of health status or care setting [4]; (2) investigated CGWT, as defined above, rather than custom-built research prototypes or specialized medical-grade hardware only [5]; (3) applied ML or AI methods to wearable-derived data for monitoring, classification, prediction, or decision support [6]; and (4) reported outcomes, tasks, or study targets related to activities of daily living, autonomy, mobility, fall detection or fall risk, frailty, functional monitoring, usability, acceptability, validation performance, or related health-support applications in older adults [4-10].

We included full primary studies published in journals or conference proceedings that provided sufficient methodological detail and results to enable eligibility assessment, data extraction, and methodological quality appraisal. Reviews, editorials, letters, study protocols, dissertations, abstracts without full study data, and other nonprimary or insufficiently detailed reports were excluded [11]. Studies were also excluded if they focused on nonolder populations, relied exclusively on custom-built experimental systems or specialized nonconsumer hardware without a CGWT, did not include an ML/AI component, or did not address outcomes relevant to autonomy, daily functioning, mobility, or fall-related applications [12].

Information Sources and Search Strategy

The search was designed and is reported in accordance with the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension) to improve transparency and reproducibility of the identification process [11]. We searched PubMed/MEDLINE, Scopus, Institute of Electrical and Electronics Engineers (IEEE) Xplore, and the Association for Computing Machinery (ACM) Digital Library from database inception to May 2026, without date or language restrictions. ScienceDirect was not searched separately because its journal content is substantially represented within Scopus, which was included as the main multidisciplinary source for Elsevier-indexed literature [13].

To reduce the risk of missing relevant studies, the search strategy combined controlled vocabulary, when supported by the database, with free-text synonyms, spelling variants,

truncation, and field-specific terms related to 4 core concepts: older adults, consumer-grade wearable or mobile sensing technologies, AI or ML methods, and autonomy-, activity-, mobility-, or fall-related outcomes and study targets [12].

Database-specific syntax, truncation, phrase searching, and field modifiers were adapted for each information source. Because a formal backward and forward citation-searching procedure was not prespecified in the PROSPERO protocol, citation searching was not treated as an additional formal identification source equivalent to the database searches.

Searches were run separately in each database rather than through a simultaneous multidatabase search platform. No study registries, manufacturer websites, or additional gray literature sources were searched. No authors, experts, or manufacturers were contacted for additional studies or unpublished data. No randomized controlled trial or other study-design filter was applied because eligible evidence was expected to include observational, methodological, feasibility, validation, system-development, and experimental studies [8,11]. Records retrieved from the databases were deduplicated before screening, and the deduplication process and database-specific retrieval counts are documented in Multimedia Appendix 1.

The search strings were developed iteratively and refined to capture broader terminology used across health, engineering, and computing literatures [13]. In addition to terms directly referring to wearable devices and ML, the revised strategy incorporated related expressions for smart devices, activity recognition, mobility, fall detection, frailty, and daily functioning, where relevant to the review question. Search execution, deduplication, and transfer to screening software [8,14] are documented in the review flow and Multimedia Appendices.

Selection and Data Collection Process

The selection of studies for this review followed a structured, multiphase process. Records retrieved from PubMed, Scopus, IEEE Xplore, and the ACM Digital Library were first exported and combined in Rayyan (Rayyan Systems, Inc) for deduplication. Duplicate records were identified using Rayyan's duplicate-detection function and manually verified by comparing bibliographic fields, including title, authors, year, journal or conference source, and DOI when available. After deduplication, the unique records were retained in Rayyan for blinded and collaborative screening [14]. The screening process consisted of 2 stages: title and abstract screening, followed by a full-text review. Both phases were conducted independently by at least 2 reviewers, who were blinded to avoid bias. In cases of disagreement, a designated senior reviewer (FAB) served as an adjudicator to ensure consensus and consistency in inclusion decisions [8].

Once the final set of included studies was determined, data extraction was performed independently by at least 2 reviewers in a double-blind fashion, consistent with SR methods intended to reduce selection and extraction errors [8,9,11]. A standardized, piloted extraction form was used to ensure consistency and reproducibility among reviewers. The extraction framework captured the following domains: bibliographic information (authors, year, and country); study characteristics (design, setting, sample size, and duration, when reported); participant characteristics; wearable device type, sensor modality, and body placement; ML/AI task and model type; comparator or reference standard, when applicable; reported outcomes; validation strategy; implementation setting (eg, laboratory, clinical, home-based, or free-living); and the study's main findings relevant to autonomy, daily functioning, mobility, fall-related applications, or related health-support outcomes [10].

For studies reporting more than one analytical task, target domain, or outcome category, extraction was conducted at 2 levels to support transparent grouping for synthesis without meta-analysis (SWiM) [15]. First, each article was retained as a single included study for purposes of study counting and the overall sample description. Second, within-study information was extracted separately for each relevant task or domain (eg, activities of daily living [ADL] monitoring, fall detection, fall-risk assessment, frailty assessment, or disease-specific monitoring) to preserve the granularity of the evidence and avoid conflating distinct objectives within the same paper. Accordingly, the same study could contribute to more than one thematic synthesis domain, but it was counted only once in the overall number of included studies.

When multiple models or performance metrics were reported within a single study, we extracted the reported approaches and summarized them narratively rather than selecting a single model as representative. Both technical outcomes (eg, accuracy, precision-recall, F_1-score, area under the receiver operating characteristic curve [AUROC], or regression error metrics) and nontechnical outcomes (eg, functional, clinical, patient-centered, usability-related, acceptability-related, or service-level outcomes) were recorded when available. For outcome-reporting classification, each included article was categorized once as either reporting only technical model-level performance metrics or reporting at least one nontechnical outcome. If a study reported both technical metrics and at least one nontechnical outcome, it was classified as having a nontechnical outcome present. This classification was recorded in the extraction table and used to summarize outcome-reporting patterns across the included studies.

Synthesis Methods

The synthesis was conducted as a structured narrative SWiM, following guidance for synthesis when statistical pooling is not appropriate [13,15]. A quantitative meta-analysis was not considered appropriate because the included studies differed substantially in several synthesis-critical dimensions, including population profiles, clinical and functional targets, wearable device types, sensor placement, analytic tasks (eg, activity recognition, fall detection, fall-risk classification, frailty screening, and disease-specific monitoring), outcome definitions, reported performance measures, and validation strategies [7-9].

In particular, many studies reported technical discrimination or prediction metrics for classification or regression tasks rather than comparable effect estimates addressing a common causal, diagnostic, prognostic, or implementation end point or prognostic association [11]. In accordance, statistical pooling would have risked combining methodologically nonequivalent evidence and generating misleading summary estimates. Instead, findings were therefore grouped by thematic application domain and interpreted with attention to methodological quality, validation approach, and clinical or functional relevance [8,15].

To improve transparency of reporting, the narrative synthesis was structured in accordance with the SWiM framework, including an explicit description of the grouping logic, synthesis methods, presentation of study characteristics and findings, and limitations affecting cross-study comparability. Studies were grouped primarily by thematic application domain and then interpreted according to setting, sensing modality, AI task, and validation approach [15]. The descriptive presentation of study characteristics and findings was therefore a deliberate methodological choice. It reflected the heterogeneity of the evidence base and the absence of comparable intervention effectiveness estimates rather than a scoping review objective.

We also categorized the studies by their implementation context and by the primary analytical task or thematic application domain reported in each study, such as ADL monitoring, gait or mobility assessment, fall detection, fall-risk assessment, frailty assessment, or disease-specific monitoring [4]. For studies addressing multiple tasks or domains, each relevant component was incorporated into the corresponding thematic synthesis, while maintaining the article as the unit of inclusion. Particular attention was paid to whether wearable-ML systems were assessed in real-world conditions or under controlled experimental setups. In this sense, each study was assessed not only for technical performance metrics but also, when available, for nontechnical outcomes relevant to interpretation and implementation [13]. These included functional or clinical outcomes, user-centered outcomes (eg, usability, acceptability, engagement, or perceived usefulness), and service-level implications. Outcome-reporting categories defined during extraction were used descriptively during synthesis [15].

Given the substantial heterogeneity in study designs, populations, sensing modalities, and AI approaches, no single quantitative effect measure was prespecified. Instead, we narratively synthesized the performance metrics reported in individual studies (eg, accuracy, precision-recall, F_1-score, and AUROC), along with any available user-centered or clinical outcomes [6-8,15].

Heterogeneity was explored descriptively by grouping studies according to target domain, sensing modality, AI task, and evaluation approach. Sensitivity analyses were not conducted because no quantitative meta-analysis was performed. To support the interpretability of the narrative synthesis, the methodological quality of the studies was considered when discussing the strengths and limitations of the evidence across domains; however, no formal certainty rating framework was applied [16].

Methodological Quality and Risk-of-Bias Assessment

Assessing methodological quality and risk of bias helps support a more cautious interpretation of the synthesized evidence. Each included study underwent an independent methodological quality assessment by 2 reviewers (IPC and FAF), with disagreements resolved through discussion or, when necessary, by a third reviewer (FAB), acting as arbiter.

We used the Prediction Model Risk-of-Bias Assessment Tool-AI (PROBAST-AI) as the primary structured tool because a substantial proportion of the included literature involved ML or AI models applied to classification, prediction, risk stratification, or related model-based decision tasks. Accordingly, the tool was considered useful for evaluating recurring concerns related to participant selection, predictors, outcome definition, model development, validation procedures, and analysis methods [16].

However, we acknowledge that the included evidence base was methodologically mixed and that not all studies aligned perfectly with conventional prediction-model designs. Some studies were closer to activity recognition, feasibility assessment, or system development than to formal prognostic or diagnostic modeling. For this reason, PROBAST-AI was applied as a structured framework to assess the risk of bias and applicability where relevant. At the same time, its limitations in noncanonical prediction-model studies were explicitly considered during interpretation. Judgments were therefore interpreted cautiously, particularly for studies with limited alignment with formal model development and validation paradigms. As is well established, small sample sizes may compromise the robustness of findings and increase the risk of bias [17].

Presentation of Results

The review results are presented in accordance with PRISMA 2020, and the search component is reported in accordance with PRISMA-S. A flow diagram outlines the full screening and selection process, including records identified through database searching, duplicates removed, records screened, reports assessed for eligibility, and studies included. The full search strategies, study characteristics, and outcome-reporting classification, PRISMA 2020 checklist, PROBAST-AI assessment, PRISMA-S checklist, and SWiM checklist are provided in Multimedia Appendices 1–4 and Checklists 1–3 [9,11,15].

In accordance with a structured SWiM approach, the results are organized in predefined thematic groupings to improve interpretability across a heterogeneous evidence base [15]. First, we present the study population, setting, and general characteristics. Second, we synthesize findings by thematic application domain, including ADL monitoring, mobility and gait assessment, fall detection or

fall-risk assessment, frailty, and disease-specific monitoring. Third, we summarize cross-cutting technological characteristics, including wearable technologies, sensor placement, ML approaches, implementation setting, validation strategy, outcome types, and methodological quality. This structure was adopted to reduce conceptual overlap and separate clinical/functional targets from technological and methodological dimensions of the evidence.

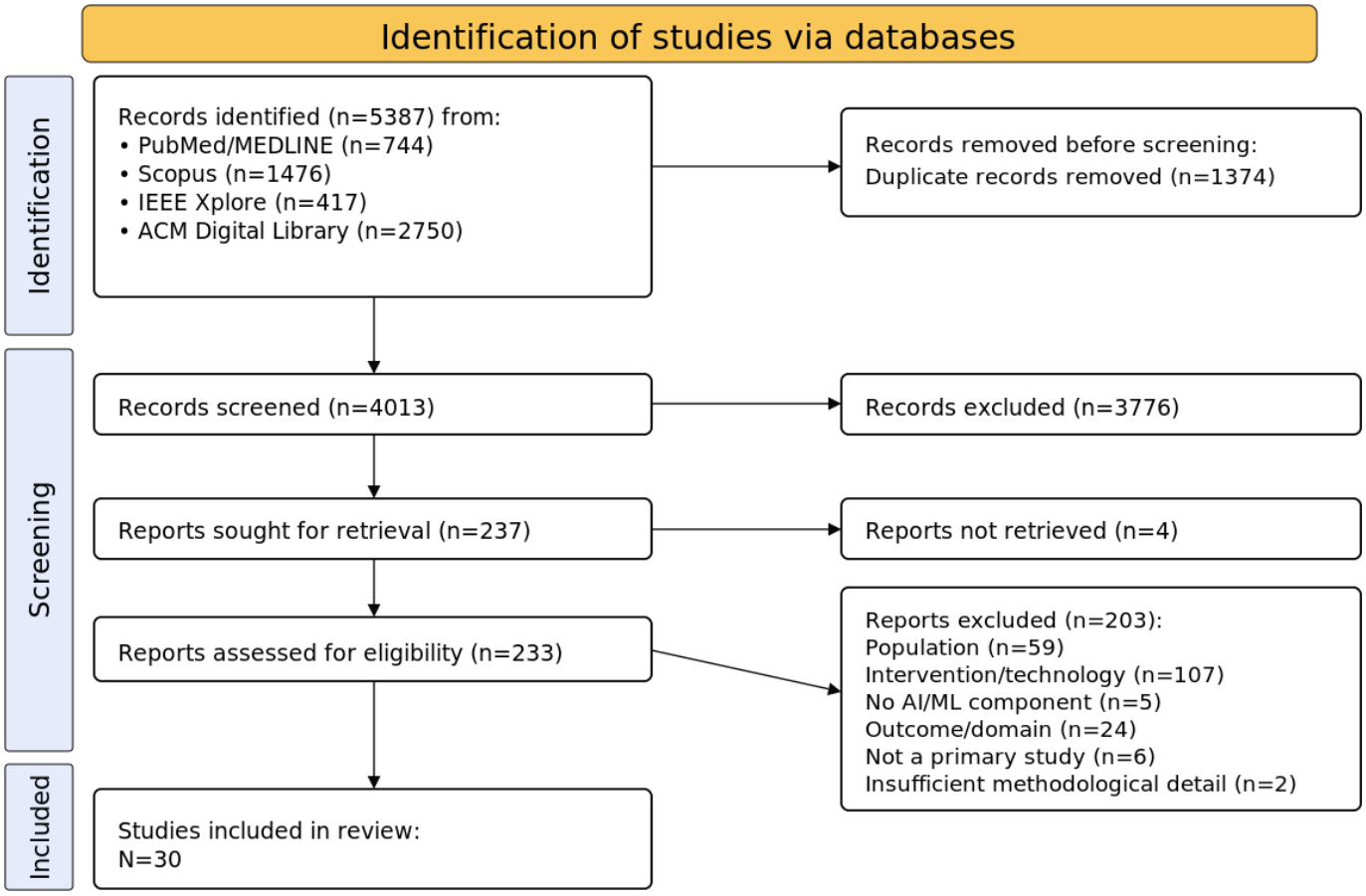
As some included studies addressed more than one application domain, these categories are not mutually exclusive; however, each study was counted only once in the total number of included studies. Findings are reported descriptively rather than statistically pooled, with emphasis on study characteristics, reported technical performance, methodological quality, and the distinction between technical performance and clinically or functionally meaningful outcomes [8]. Areas of inconsistency, limited external validation, and reduced applicability to real-world care are explicitly highlighted.

Results

Study Selection

The systematic search yielded 5387 records. After removing 1374 duplicates, 4013 records were screened by title and abstract, and 3776 were excluded. A total of 237 reports were sought for retrieval, of which 4 could not be retrieved. Therefore, 233 reports were assessed for eligibility. After full-text assessment, 203 articles were excluded due to population (n=59), intervention/technology (n=107), no AI/ML component (n=5), outcome/domain (n=24), not a primary study (n=6), or insufficient methodological detail (n=2). In the screening process, most title/abstract exclusions were due to irrelevance to older adult populations, absence of wearable technologies or ML components, or nonoriginal study designs. In the full-text assessment, exclusions were mainly due to an ineligible population, intervention, or outcome, including a lack of AI-based analysis, absence of relevant wearable data, or insufficient methodological detail. Finally, 30 studies [18-47] were included in this SR (Figure 1).

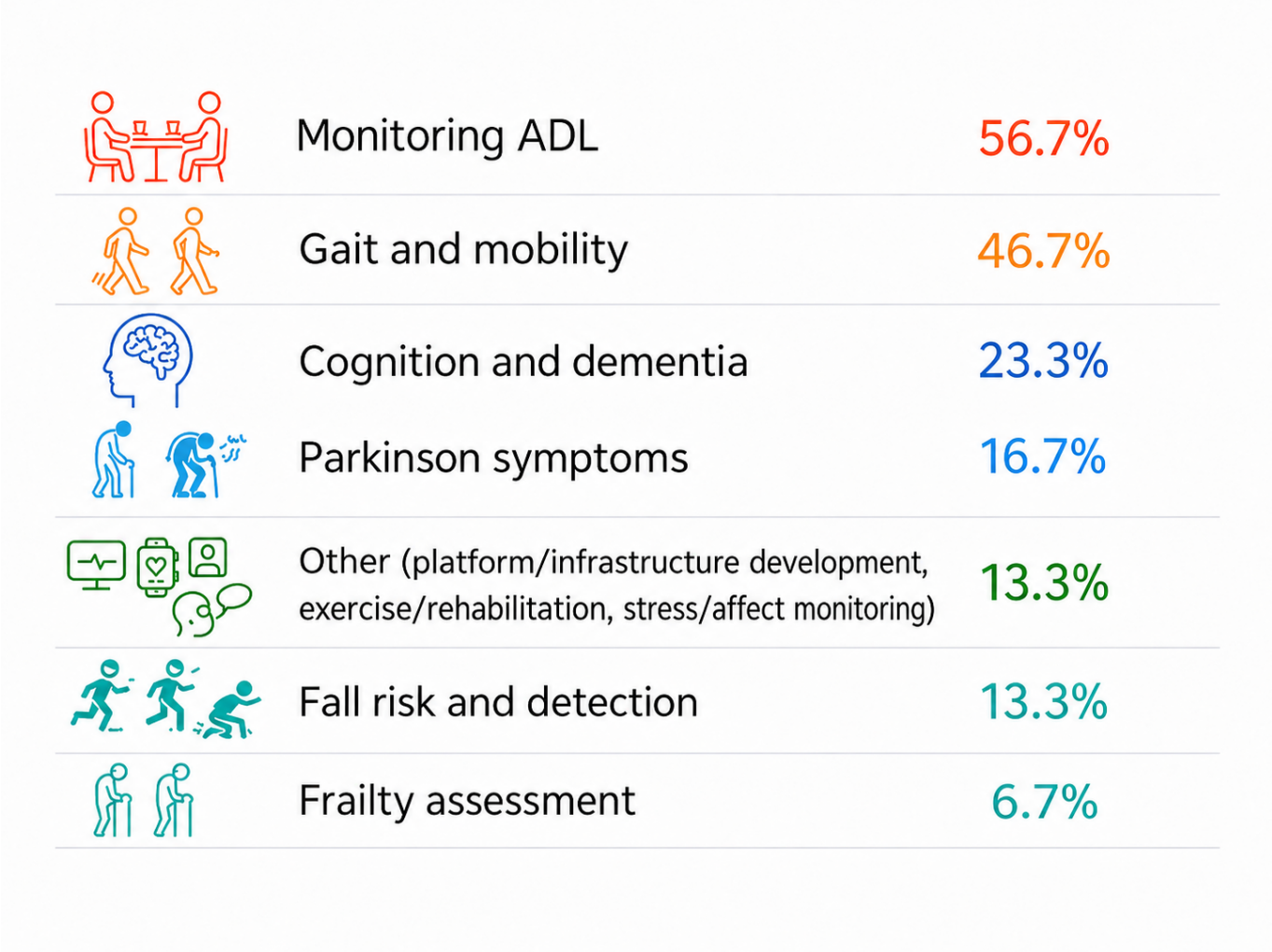
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram of study selection, including records identified through database searching, duplicate records removed, records screened, full-text reports assessed for eligibility, and studies included in the systematic review. ACM: Association for Computing Machinery; IEEE: Institute of Electrical and Electronics Engineers; ML: machine learning.



Study Characteristics and Thematic Application Domains

The included studies addressed several thematic application domains related to wearable and ML-based support for older adults. These domains included ADL, gait and mobility assessment, cognitive- or dementia-related functional monitoring, Parkinson-related monitoring, fall detection or fall-risk assessment, and frailty assessment. Several studies addressed more than one domain simultaneously. In such cases, the study was retained as a single included article but contributed to each relevant thematic category during synthesis. Therefore, the categories summarized in Figure 2 are not mutually exclusive and should be interpreted as within-study domain contributions rather than independent study counts. This thematic structuring also supported interpretation across a heterogeneous evidence base. Across domains, the literature was unevenly distributed, with ADL and mobility-related monitoring among the most frequently represented areas, whereas frailty and disease-specific applications were studied less consistently. The most frequent thematic domains were ADL monitoring or activity recognition (17/30, 56.7%) [18,19,21,22,24-28,30,32,33,35,38-41] and gait or mobility assessment (14/30, 46.7%) [19,22,24,31,33,34,36,37,39,40,42,43,46,47], followed by cognitive/dementia-related monitoring (7/30, 23.3%) [18-20,25,42,46,47], Parkinson-related monitoring (5/30, 16.7%) [21,29,43-45], fall detection or fall-risk assessment (4/30, 13.3%) [31,33,34,37], frailty assessment (2/30, 6.7%) [20,36], and other platform, implementation, exercise/rehabilitation, or stress/affect monitoring applications (4/30, 13.3%) [20,23,30,39]. Because categories were not mutually exclusive, percentages may sum up to more than 100%.

Figure 2. Distribution of thematic application domains across the 30 included studies [18-47] in this systematic review.



Population, Setting, and Implementation Context

This section summarizes the main characteristics of the study populations included in the review. The included studies varied in sample size, participant age, and clinical characteristics, with most focusing on older adults, including individuals with Parkinson disease, cognitive impairment, or increased risk of frailty. Studies were conducted in diverse settings, such as community-based environments, clinical settings, and long-term care facilities. An overview of the key population and study characteristics of the 30 included studies is provided in Table S1 in Multimedia Appendix 2. This appendix also includes the study-level outcome-reporting

classification used to distinguish studies that report only technical model performance metrics from those that report at least one functional, clinical, patient-centered, usability-related, or service-level outcome.

The sample sizes of the included studies were highly heterogeneous. Studies involving direct participant recruitment typically included small samples, often with fewer than 20 participants, particularly in exploratory or pilot research [18-21]. In contrast, a small number of studies relied on large, database-driven or multicenter datasets, with sample sizes exceeding 500 individuals [22,23].

Regarding age distribution, studies that used direct participant recruitment included only participants aged 60 years or older, whereas studies based solely on simulated data or artificial falls excluded participants aged 60 years or older. Participants were therefore predominantly frail or healthy older adults. Some studies reported populations with Parkinson disease, dementia, or multimorbidities, particularly cardiovascular conditions, reflecting the increasing prevalence of degenerative diseases among older adults [3,4].

Overall, the studies included older adults with moderate functional capacity and sufficient mobility to perform the proposed tasks. Most participants were female, reflecting not only higher average life expectancy than men [48] but also women's greater availability and receptivity to participating in studies of this nature [49].

Moreover, the included studies showed marked heterogeneity in study design, setting, and duration. Most evaluations were conducted in home-based or free-living (real-world) settings (19/30, 63.3%) [18,19,22-25,28,30,31,33,35,36,38-42,46,47], followed by clinical or institutional settings (8/30, 26.7%) [20,26,29,34,37,43-45], laboratory or simulated environments (2/30, 6.7%) [21,27], and mixed or unclear contexts (1/30, 3.3%) [32]. Across these settings, study designs were predominantly observational or methodological (often feasibility testing or algorithm validation), whereas experimental or randomized designs were uncommon. Study durations were generally short and heterogeneous, with most studies relying on brief protocols and few longitudinal designs with extended follow-up. This pattern reflects the early-stage nature of the current evidence base and limits its capacity to capture long-term outcomes in older adults. Limited stakeholder involvement may reduce the clinical interpretability and implementation relevance of proposed systems.

For example, 11 studies [18,19,21,23-30] described the participation of health care stakeholders (eg, nursing professors, gerontologists, caregivers, and clinical assessors) in recruitment and/or data interpretation, or reported the engagement of multidisciplinary clinical teams. In contrast, the remaining studies either did not report stakeholder involvement [20,22,31-37,42-45,47] or explicitly stated the absence of patient, caregiver, or health care professional co-design [38-41]. This highlights a substantial gap in multidisciplinary collaboration in the development and validation of AI systems for older adults.

Results by Thematic Application Domain

The thematic synthesis grouped studies according to their primary functional application rather than treating wearable-based ML as a homogeneous evidence base. Five main areas were identified: ADL, gait and mobility assessment, fall detection and fall-risk assessment, frailty assessment, and disease-specific functional monitoring. Within these domains, smartphone- and smartphone-plus-wearable studies contributed mainly to mobility-related, cognitive/dementia-related, and Parkinson-related monitoring. These studies included daily-life gait sensing, multidomain smartphone assessment, smartphone-based motor assessment, touchscreen-based fine-motor tasks, and passive mobile sensing of mobility or cognitive status [42-47].

In the ADL domain, studies focused on recognizing routine behaviors, instrumental tasks, exercise-related activities, or home-based activity patterns using wearable inertial and, less commonly, multimodal signals. These studies primarily emphasized activity recognition performance and feasibility, with variable attention to ecological validity and limited reporting of downstream functional or care-related implications. This domain was strongly represented and often relied on inertial sensing; smartphone- and smartphone-plus-wearable studies also addressed daily-life gait, mobility signatures, walking, or postural-stability–related features [42-44,46,47]. However, comparability remained limited by differences in sensor placement, task definitions, outcome labels, and evaluation settings.

In the fall detection and fall-risk assessment domain, studies addressed either real-time/event-level detection or classification of individuals according to fall risk. Although this domain is highly relevant to autonomy and safety, studies differed substantially in whether they used simulated tasks, laboratory protocols, or free-living data, which limits cross-study interpretation and real-world inference. In the frailty assessment domain, the literature was smaller and generally exploratory, with wearable-derived measures used to characterize physical vulnerability or functional decline. These studies suggest potential relevance but remain limited in number and methodological consistency.

In the disease-specific functional monitoring domain, studies focused mainly on Parkinson-related monitoring or cognitive/dementia-relevant functional assessment. Smartphone- and mobile-sensing approaches were particularly prominent in this domain, including studies of cognitive impairment or cognitive status [42,46,47] and Parkinson-related motor assessment or early identification [43-45]. These studies illustrate the potential of CGWT and ML in specific clinical contexts, but they should be interpreted separately from broader autonomy-support applications because the target constructs, reference standards, validation strategies, and intended uses are not identical. Most studies in this subgroup primarily reported model-performance outcomes rather than demonstrated clinical effectiveness or service-level impact. Overall, this domain-based structure shows that the evidence base remains fragmented across functional constructs, reference standards, data-collection contexts, and

intended clinical uses. Therefore, findings from specific domains should not be generalized uncritically to broader claims about autonomy support in older adults.

Wearable Technologies, Sensor Placement, and ML Approaches

This subsection summarizes cross-cutting technological characteristics across the included studies and should be interpreted separately from the thematic application domains presented above. The characteristics of wearable technology and the reported ML approaches are summarized in Multimedia Appendix 4. Because categories may overlap across studies, counts should be interpreted as reporting frequency rather than mutually exclusive groups.

Overall, no study referenced the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis-Artificial Intelligence (TRIPOD-AI) checklist [50]; instead, most papers described their approach primarily by task type. Overall, most included studies framed AI applications primarily as classification tasks. Because some studies reported more than one analytical task or model family, algorithm and task categories were treated as reporting frequencies rather than mutually exclusive study groups. The most frequently reported model families were random forests (13/30) [18-20,23,26,36-40,43,44,4618-20], support vector machines (12/30) [18-20,26,32,34,36-38,43,44,46], convolutional neural networks (9/30) [18,19,22,29,31,35,39,45,46], long short-term memory models or related recurrent architectures (9/30) [29,31,35,39,42,44-47], k-nearest neighbors (7/30) [18-20,26,32,44,46], decision trees (5/30) [20,32,33,37,44], and ConvLSTM (1/30) [31]. Sensor placement, sensing modality, and algorithm categories may overlap across studies.

With respect to privacy and ethical safeguards, 9 studies [19,24,26,27,29,32-35] did not report ethics approval and/or data protection procedures. In contrast, 17 studies [20,21,25,30,33,36-47] reported ethics approval (eg, ethics committee approval and/or informed consent), and 11 studies [18,20,22,23,30,36-38,41,46,47] described at least one data protection measure (eg, anonymization/deidentification, secure storage, or restricted data sharing).

Most studies reported monitored variables and input features, predominantly derived from accelerometry. A smaller subset of studies used physiological or barometric data to improve the interpretation of activity, mobility, or fall-related events.

Validation Strategies, Reported Outcomes, and Risk of Bias

This subsection presents cross-cutting evaluation characteristics across domains, including validation strategy, reported outcome type, and methodological quality, rather than domain-specific findings. Evaluation strategies and performance reporting varied substantially across the included studies. Consistent with this predominance of classification tasks, most studies reported discrimination-focused metrics, including accuracy, sensitivity/recall, specificity, precision, and F_1-score, sometimes complemented by confusion matrices and, less frequently, AUC-based measures. Studies addressing regression outcomes generally relied on error-based metrics (eg, mean absolute error, mean squared error, and root mean squared error) or correlation-based measures. However, reporting was heterogeneous and often lacked uncertainty estimates, limiting cross-study comparability.

When outcomes were categorized according to their interpretive level using the study-level classification documented in Table S2 in Multimedia Appendix 2, 15/30 (50%) studies [21,26,27,31,32,34,35,37-39,41-43,45,47] reported only technical model performance metrics, such as accuracy, sensitivity/recall, specificity, precision, F_1-score, AUROC, or regression error measures, without reporting functional, clinical, patient-centered, usability-related, or service-level outcomes. In contrast, 15/30 (50%) [18-20,22-25,28-30,33,36,40,44,46] studies reported at least one nontechnical outcome, such as usability, acceptability, functional relevance, clinical interpretation, patient-centered assessment, or potential service implications. This distribution reinforces that the current evidence base is more mature in technical evaluation than in demonstrating clinical usefulness or implementation readiness.

Because the included literature was methodologically diverse, PROBAST-AI findings should be interpreted primarily as a structured appraisal of model-related risk-of-bias and applicability concerns rather than as a perfectly uniform assessment across identical study designs. Using PROBAST-AI, we assessed risk of bias across the 4 standard domains (participants, predictors, outcome, and analysis) for the included prediction-model studies. Most studies were classified as combination studies (model development and evaluation), while a smaller subset reported development only, and one study [40] focused primarily on model evaluation. Among the studies with sufficient information to judge model evaluation (n=25) [19-22,24-27,29-39,42-47], the predictors domain showed the highest methodological quality, with low risk of bias across all assessments (25/25) [19-22,24-27,29-39,42-47]. The outcome domain was also largely adequate (23/25 low risk) [19-22,25-27,29-36,38,39,42-47], with only a minority raising concerns related to outcome definition and/or determination.

In contrast, limitations were more frequent in the participants' domain (11/25 low risk [19,22,24,26,27,29-31,33,36,44], 10/25 high risk [21,25,34,35,37,42,43,45-47], and 4/25 unclear) [20,32,38,39], typically reflecting heterogeneous or poorly described data sources, limited representativeness, or incomplete reporting of inclusion criteria and recruitment procedures. The analysis domain concentrated the main concerns (9/25 low risk [19,20,22,26,31,34,39,43,44], 12/25 high risk [21,24,25,27,29,30,35,37,42,45-47], and 4/25 unclear [32,33,36,38]), consistent with incomplete reporting and suboptimal evaluation practices. Specifically, PROBAST-AI concerns included frequent reliance on internal cross-validation without external validation, potential leakage between feature selection and validation folds, limited

reporting of calibration, and sparse handling of missingness or wear time. Variability in outcome definitions between scripted laboratory tasks and proxy free-living labels also affected applicability [16].

Regarding model evaluation approaches, external or independent validation remained limited. Parkinson-related studies included examples of external validation and independent testing, but most studies still relied mainly on internal cross-validation, apparent performance, or single-site testing. This remains a major limitation because it weakens confidence in model transportability across populations, settings, devices, and care contexts. Per-study domain judgments are provided in Multimedia Appendix 3.

Discussion

Principal Findings

This SR synthesized heterogeneous evidence from 30 studies [18-47] on CGWT combined with AI/ML in older adults across autonomy-relevant domains, including ADL, gait and mobility assessment, fall detection or fall-risk assessment, frailty assessment, and disease-specific functional monitoring. Given the diversity of aims, outcomes, and validation strategies, the findings are best interpreted by thematic domain rather than as a single unified outcome area. Across the included studies, inertial data remained the predominant sensing modality, while smartphone- and mobile-sensing approaches also contributed to mobility, cognitive/dementia-related, and Parkinson-related monitoring [42-47]. Most studies were observational or methodological, focusing on feasibility, activity recognition, model validation, or risk stratification rather than causal evaluation of clinical benefit [20,27,33].

A major interpretive constraint was the substantial heterogeneity of participant profiles, clinical contexts, wearable platforms, sensor placement, analytic tasks, outcome definitions, and validation methods, which limited comparability and confidence in broad real-world inferences [8,16,50]. In practical terms, strong performance reported for one task, device configuration, or study population cannot be assumed to transfer to other older-adult groups, care contexts, or implementation settings without further validation.

Domain-specific interpretation is also important. Evidence related to ADL and mobility monitoring was broader but remained dominated by technical performance and feasibility-oriented evaluation [25,32,39]. Fall-related applications were clinically salient but methodologically mixed, particularly with respect to simulated versus free-living data [31,33,36]. Frailty and disease-specific monitoring appeared promising but were supported by smaller and less consistent evidence bases [18-21,37]. This pattern cautions against broad claims about "autonomy support" without specifying the domain, task, and validation context.

Smartphone- and mobile-sensing approaches contributed to cognitive/dementia-related and Parkinson-related functional monitoring, including daily-life gait sensing, multidomain smartphone assessment, smartphone-based motor assessment, touchscreen-based fine-motor tasks, and passive mobile sensing [42-47]. These studies broadened the technological scope of the review toward mobile and passive sensing, while reinforcing the main interpretive limitation that the evidence supports technical feasibility more clearly than clinical usefulness, implementation readiness, or patient-centered benefit. Limited stakeholder involvement may further reduce the clinical interpretability and implementation relevance of proposed systems [51-53].

Therefore, there is a need for more robust trial designs and extended follow-up periods to enhance clinical relevance, as recently emphasized in the TRIPOD-AI guidelines [50]. Although the results show that none of the included studies explicitly reported adherence to TRIPOD-AI, adopting structured reporting frameworks is likely to improve transparency, reproducibility, and comparability in prediction-model research for older adults, facilitating external validation and clinical translation.

The combination of wearable technologies and ML may support earlier recognition of functional decline, fall-related vulnerability, and patterns relevant to individualized care in older adults. However, the current review indicates that these possibilities remain more strongly supported at the level of technical development than at the level of demonstrated clinical usefulness [50,54,55]. Accordingly, potential contributions to independence, continuity of care, or health care efficiency should be interpreted cautiously until supported by stronger real-world and patient-centered evidence [1,4].

A central finding is the tension between technical accuracy and clinical actionability. Many studies demonstrated that wearable- or smartphone-derived models could recognize activities, detect mobility patterns, or classify disease-related functional signals with encouraging model-level performance [20,27,33,42-47]. This distinction was also evident in the outcome-reporting pattern: 15/30 (50%) [20,22-25,28-30,33,36,40,44,46] studies reported only technical performance metrics, whereas 15/30 (50%) [21,26,27,31,32,34,35,37-39,41-43,45,47] reported at least one functional, clinical, patient-centered, usability-related, or service-level outcome. Still, translation into decision support for prevention, rehabilitation, or community care was less frequently observed.

Work on real-world, free-living movement patterns and bout-level gait or posture inference illustrates both the opportunity and the challenge: algorithms trained on laboratory or structured tasks often require adaptation, calibration, or retraining to remain robust in everyday settings with short, irregular activity segments and variable device positioning [22,38]. Studies that collected or tested data under free-living conditions reported more nuanced performance, underscoring the importance of ecological validity during development and evaluation [21,27,33].

A second cross-cutting theme concerns sensor placement and signal sufficiency. Wrist devices dominate because of comfort and availability, yet gait stability, turns, and

near-falls can be more salient at the lower back or waist; several studies exploited trunk or hip placement to capture center-of-mass dynamics and transitional movements, whereas others pursued fusion of wrist inertial measurement unit (IMU) with HR/photoplethysmography or skin conductance to contextualize exertion, stress, or arousal during daily activities [26,31,34,36,37]. Collectively, these results suggest that modality and placement should be chosen according to the intended decision: screening for general activity levels may tolerate wrist-only signals, while fall-risk phenotyping, transitional instability, or rehabilitation progression may benefit from waist/back or multimodal configurations [18,24,25].

In addition, IMU-based sensing predominated (26/30 studies) [18,19,21-29,31-42,44,46,47], reflecting the substantial emphasis on gait and fall-related monitoring in older adults (Multimedia Appendix 4). However, smartphone/mobile-sensing approaches formed a clearer subgroup, including touchscreen, voice, video, GPS/location, app use, typing, and smartphone-smartwatch gait features.

Data transmission strategies were frequently underreported: 14 of 30 studies [18,19,21,22,24,25,31-35,40,43,45] did not describe how data were transferred from wearables to smartphones or computers (Multimedia Appendix 4). When described, periodic data off-loading/batch upload (8 studies) [36,38,39,41,42,44,46,47] may limit real-time detection workflows [44], while Bluetooth transmission (5 studies) [20,23,26,28,30] requires proximity between the wearable and the receiving device. Only a small number of studies reported other approaches (eg, wireless transmission without protocol specification or connections via microcontrollers), underscoring implementation gaps that affect real-world feasibility.

Finally, ethical and privacy safeguards were inconsistently reported. In the results, 9 studies [19,24,26,27,29,32,35] did not report ethics approval and/or data protection procedures, highlighting the need for clearer statements on consent, data handling, and privacy protections to support the acceptability and deployment of remote monitoring for older adults.

Third, model families span traditional ML pipelines and deep learning architectures. Classical approaches, such as support vector machines, decision trees, random forests, and k-nearest neighbors, remain prevalent due to their transparency, modest data requirements, and on-device feasibility [27,28,32]. At the same time, convolutional and recurrent networks, hybrids, and sequence-aware models have been applied to human activity recognition and risk modeling, leveraging raw or minimally engineered signals and enabling end-to-end learning [31,35,36].

Recent studies also explore the hierarchical recognition of structured exercise programs at home, illustrating how model outputs can be organized into clinically meaningful units, such as sessions and components of rehabilitation protocols. This thereby reduces reliance on self-report and enables objective adherence tracking [29,40]. These shifts align with the needs of digital rehabilitation and proactive care, where timely, interpretable summaries often take precedence over frame-level classification accuracy [39].

Methodological quality varied. Samples were frequently small and convenience-based, with short observation windows that limited inferences about durability, seasonality, and device fatigue [24,28,33]. Ground-truth definitions and labeling procedures were heterogeneous, ranging from scripted tasks to proxy labels in the wild, which complicated cross-study comparability and precluded meta-analysis [20,21,34]. External or independent validation remained limited; although some disease-specific studies included external validation or independent testing, most studies relied mainly on internal cross-validation, apparent performance, or single-site testing [42-47]. This limited interpretation of implementation readiness and transportability across device brands, firmware versions, sampling behavior, body placement, population characteristics, and care contexts.

PROBAST-AI provided a useful structured lens for recurring model-related concerns, particularly around participant selection, analysis, and validation. However, because some included studies were closer to activity recognition, feasibility assessment, or system development than to formal prediction modeling, these judgments should be interpreted as structured model-related appraisals rather than uniform assessments of identical study designs [16].

The choice of a structured eligibility framework rather than an effectiveness-oriented population, intervention, comparison, and outcome framework was deliberate [8,10]. The objective was not to compare one intervention against another or to estimate pooled effects on predefined clinical end points. Instead, the review aimed to systematically identify, appraise, and synthesize primary studies on CGWT combined with ML in older adults, with attention to application domains, methodological characteristics, validation strategies, and reported outcomes [8,9]. Given the heterogeneity of populations, devices, analytical tasks, validation strategies, and outcome definitions, this approach supported a structured narrative synthesis rather than statistical pooling [15].

From an implementation standpoint, several studies suggested possible service-integration pathways, including free-living activity recognition for adherence monitoring, unobtrusive sensing of cognitive or instrumental ADLs, privacy-preserving interaction modalities, and smartphone-based workflows involving telemedicine, General Data Protection Regulation (GDPR)–oriented data handling, or user-centered development [22,26,29,30,36,39,40,46]. However, these examples should be interpreted as early implementation signals rather than evidence of routine clinical readiness, because sustained usability, workflow integration, privacy safeguards, and service-level impact remain insufficiently evaluated.

Beyond model performance, implementation of wearable- and AI-based systems in older adults requires attention to ethical, legal, operational, and user-centered factors. Continuous data collection raises privacy and data-protection concerns, requiring transparent communication with

users, secure data handling, and compliance with regulatory frameworks such as the GDPR [56]. Digital health literacy may also affect adoption, because older adults vary in their ability to access, understand, and use digital health technologies [57,58]. In addition, the absence of clear regulatory frameworks and medical-grade validation for many consumer devices remains a barrier to safe clinical integration [59]. Sustained use also depends on comfort; ease of use, perceived usefulness; accessibility for people with sensory, motor, or cognitive impairments; and co-design with older adults, caregivers, and clinicians [51-53]. These issues remain unevenly addressed and should be evaluated alongside model performance before claims about real-world implementation can be justified.

The main contribution of this review is to integrate evidence across application domains that are often discussed separately and to clarify the gap between technical performance and clinically meaningful usefulness. The review supports cautious optimism regarding the future of the field, but the current evidence remains limited by heterogeneous methods and insufficient demonstration of real-world clinical usefulness or implementation readiness.

Limitations and Future Research

This review has limitations that warrant consideration. The included studies were highly heterogeneous in terms of populations, devices, sensing strategies, analytic tasks, outcome definitions, and validation procedures. As a result, quantitative pooling was not methodologically appropriate, and the findings were synthesized narratively. This heterogeneity also limited cross-study comparability and weakened the interpretability of aggregate claims across domains, particularly with regard to reproducibility and real-world transportability [18,20]. Selective reporting bias cannot be excluded, as reporting was often incomplete regarding sampling frames, annotation protocols, and handling of missing data, which constrained risk-of-bias assessment and external validity [21,33].

Although the database search strategy was revised and updated, this review did not include a full, formal backward and forward citation-searching procedure as a prespecified supplementary search method. We also did not search manufacturer websites, contact authors or device manufacturers, or conduct a full gray-literature search. Therefore, relevant unpublished studies, technical reports, commercial validation data, or studies identifiable only through citation chasing may not have been captured. Future updates of this review should prospectively incorporate backward and forward citation searching as a predefined supplementary identification method, with complete documentation of sources searched, dates of access, and screening procedures.

We did not formally assess reporting bias or publication bias, as no quantitative synthesis or meta-analysis was conducted, and outcome reporting varied substantially across studies. The overall certainty of evidence was not formally evaluated because the included studies were predominantly early-stage, methodologically heterogeneous, and not suitable for pooled quantitative estimation.

Nevertheless, by consolidating findings across sensing modalities, placements, and modeling strategies, the review provides a coherent view of where translation is most imminent and where methodological investment is needed [27,37]. Although PROBAST-AI offered a structured way to appraise recurring model-related concerns [16], its application to a methodologically mixed body of literature has limitations, particularly for studies oriented toward activity recognition, feasibility assessment, or system development rather than formal prediction modeling. In addition, incomplete reporting in several primary studies limited domain-level certainty, particularly regarding calibration, handling of missingness, and label provenance. Consequently, some domain ratings were “unclear,” and the resulting inferences were kept conservative.

Taken together, these observations inform priorities for future research. First, algorithm development should adopt free-living data as the default, with explicit handling of short bouts, missing data, wear-time detection, and device rotation, and with pretraining or self-supervised learning to leverage large unlabeled corpora before supervised fine-tuning [22,35]. Second, studies should report transparent pipelines, calibration, and uncertainty, and provide external validation across sites and devices to quantify real-world transportability [31,36,41,50]. Third, prospective real-world evaluations, including pragmatic trials where appropriate, are needed to examine patient-centered and service outcomes, ideally by embedding interoperable data flows into electronic records [25,29,39]. Fourth, co-design with older adults, caregivers, and clinicians should be the norm to ensure usability, acceptability, and sustained adherence, especially in populations with sensory, motor, or cognitive limitations [30,40].

Conclusion

This SR synthesized 30 studies [18-47] on consumer-grade wearable and related mobile sensing technologies combined with ML in older adults. The evidence suggests technical promise across autonomy-relevant domains, including ADL monitoring, mobility assessment, fall-related applications, frailty evaluation, and disease-specific functional monitoring. However, these domains differ substantially in aims, outcome definitions, validation strategies, and clinical maturity; therefore, findings should be interpreted by thematic domain rather than as evidence from a single homogeneous field.

Overall, the literature supports feasibility and model-level performance more clearly than patient-centered benefit, clinical usefulness, or implementation readiness. Small and heterogeneous samples, limited external validation, short follow-up, inconsistent reporting, and infrequent service-level or patient-centered outcomes remain important constraints. Future studies should prioritize free-living evaluation, transparent reporting, external validation across populations and devices, interdisciplinary co-design, and pragmatic assessment of functional, patient-centered, and service-level outcomes before stronger claims about routine clinical adoption can be justified.

Generative AI Use

The authors used generative AI as a language-support tool during manuscript preparation. All generated or AI-assisted text was reviewed, edited, and verified by the authors, who take full responsibility for the content of the manuscript.

Funding

This research was developed within the framework of the Portuguese consortium of the RM4Health project (ITEA-2021-21022-RM4HEALTH), exclusively funded by the European Union through the European Regional Development Fund (ERDF), under the Portugal 2030 framework, within the scope of the Innovation and Digital Transition Programme – COMPETE 2030 (operation no. 12949 / project no. COMPETE2030-FEDER-00391100). This article was supported by National Funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., within the project RISE-Health - UID/06397/2025 (DOI: 10.54499/UID/06397/2025). FAB, VL, and RA acknowledge support from the Portuguese Foundation for Science and Technology (FCT) through FCT-Tenure-funded positions at the Faculty of Medicine of the University of Porto (2023.13694.TENURE.043, 2023.13694.TENURE.055, and 2023.13694.TENURE.056, respectively).

Data Availability

All data used in this systematic review were extracted from published studies and are available in Multimedia Appendix 2 (study characteristics and outcome reporting classification) and Multimedia Appendix 3 (Prediction Model Risk-of-Bias Assessment Tool–AI judgments). No individual-level participant data were accessed. Researchers interested in individual-level data from the studies included in this study are encouraged to contact the corresponding author of each study for such requests.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full reproducible electronic search strategies for all databases.
[XLSX File (Microsoft Excel File), 83 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Characteristics of included studies and study-level outcome-reporting classification.
[XLSX File (Microsoft Excel File), 90 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Prediction Model Risk-of-Bias Assessment Tool–AI risk of bias assessment for all included studies.
[XLSX File (Microsoft Excel File), 151 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Summary of technological characteristics and machine learning approaches reported across the 30 included studies.
[DOCX File (Microsoft Word File), 873 KB-Multimedia Appendix 4]

Checklist 1

PRISMA 2020 checklist
[DOCX File (Microsoft Word File), 1962 KB-Checklist 1]

Checklist 2

PRISMA-S checklist.
[DOCX File (Microsoft Word File), 20 KB-Checklist 2]

Checklist 3

SWiM checklist.
[DOCX File (Microsoft Word File), 31 KB-Checklist 3]

References

1. The Lancet Regional Health – Europe. Securing the future of Europe's ageing population by 2050. Lancet Reg Health Eur. Dec 2023;35:100807. [doi: 10.1016/j.lanepe.2023.100807] [Medline: 38115962]
2. Gianfredi V, Nucci D, Pennisi F, Maggi S, Veronese N, Soysal P. Aging, longevity, and healthy aging: the public health approach. Aging Clin Exp Res. Apr 17, 2025;37(1):125. [doi: 10.1007/s40520-025-03021-8] [Medline: 40244306]
3. Martinez R, Morsch P, Soliz P, Hommes C, Ordunez P, Vega E. Life expectancy, healthy life expectancy, and burden of disease in older people in the Americas, 1990-2019: a population-based study. Rev Panam Salud Publica. 2021;45:e114. [doi: 10.26633/RPSP.2021.114] [Medline: 34621302]

4. Rudnicka E, Napierała P, Podfigurna A, Męczekalski B, Smolarczyk R, Grymowicz M. The World Health Organization (WHO) approach to healthy ageing. Maturitas. Sep 2020;139:6-11. [doi: 10.1016/j.maturitas.2020.05.018] [Medline: 32747042]
5. Shajari S, Kuruvinashetti K, Komeili A, Sundararaj U. The emergence of AI-based wearable sensors for digital health technology: a review. Sensors (Basel). Nov 29, 2023;23(23):9498. [doi: 10.3390/s23239498] [Medline: 38067871]
6. Qi W, Zhu X, Wang B, et al. Alzheimer's disease digital biomarkers multidimensional landscape and AI model scoping review. NPJ Digit Med. Jun 16, 2025;8(1):366. [doi: 10.1038/s41746-025-01640-z] [Medline: 40523935]
7. Qi W, Shen S, Dong C, et al. Digital biomarkers for Parkinson disease: bibliometric analysis and a scoping review of deep learning for freezing of gait. J Med Internet Res. May 20, 2025;27:e71560. [doi: 10.2196/71560] [Medline: 40392578]
8. Higgins JPT, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions. Wiley; 2019. [doi: 10.1002/9781119536604] ISBN: 9781119536628
9. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
10. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res. Nov 21, 2014;14:579. [doi: 10.1186/s12913-014-0579-0] [Medline: 25413154]
11. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. Jan 26, 2021;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
12. Wang S, Li H, Scells H, Locke D, Zuccon G. MeSH term suggestion for systematic review literature search. In: Mackenzie J, Spina D, editors. ADCS '21: Proceedings of the 25th Australasian Document Computing Symposium. Association for Computing Machinery; 2022. [doi: 10.1145/3503516.3503530]
13. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods. Mar 2020;11(2):181-217. [doi: 10.1002/jrsm.1378] [Medline: 31614060]
14. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. Dec 5, 2016;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
15. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. Jan 16, 2020;368:l6890. [doi: 10.1136/bmj.l6890] [Medline: 31948937]
16. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. BMJ. Mar 24, 2025;388:e082505. [doi: 10.1136/bmj-2024-082505] [Medline: 40127903]
17. Faber J, Fonseca LM. How sample size influences research outcomes. Dental Press J Orthod. 2014;19(4):27-29. [doi: 10.1590/2176-9451.19.4.027-029.ebo] [Medline: 25279518]
18. Alam MAU, Roy N, Holmes S, Gangopadhyay A, Galik E. Automated functional and behavioral health assessment of older adults with dementia. In: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). IEEE; 2016. [doi: 10.1109/CHASE.2016.16]
19. Alam MAU, Roy N, Holmes S, Gangopadhyay A, Galik E. AutoCogniSys: IoT assisted context-aware automatic cognitive health assessment. In: Mühlhäuser M, Polyzos GC, Michahelles F, Sanchez Guinea A, Wang L, editors. MobiQuitous '20: MobiQuitous 2020 - 17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. Association for Computing Machinery; 2021:184-195. [doi: 10.1145/3448891.3448945]
20. Delmastro F, Martino FD, Dolciotti C. Cognitive training and stress detection in MCI frail older people through wearable sensors and machine learning. IEEE Access. 2020;8:65573-65590. [doi: 10.1109/ACCESS.2020.2985301]
21. Nguyen H, Lebel K, Bogard S, Goubault E, Boissy P, Duval C. Using inertial sensors to automatically detect and segment activities of daily living in people with Parkinson's disease. IEEE Trans Neural Syst Rehabil Eng. Jan 2018;26(1):197-204. [doi: 10.1109/TNSRE.2017.2745418] [Medline: 28858808]
22. Brand YE, Kluge F, Palmerini L, et al. Self-supervised learning of wrist-worn daily living accelerometer data improves the automated detection of gait in older adults. Sci Rep. Sep 6, 2024;14(1):20854. [doi: 10.1038/s41598-024-71491-3] [Medline: 39242792]
23. Bellandi V. A big data infrastructure in support of healthy and independent living: a real case application. In: Lim CP, Vaidya A, Chen YW, Jain V, Jain LC, editors. Artificial Intelligence and Machine Learning for Healthcare, Vol 2: Emerging Methodologies and Trends. Springer; 2023:95-134. [doi: 10.1007/978-3-031-11170-9_5]

24. Saporito S, Brodie MA, Delbaere K, et al. Remote timed up and go evaluation from activities of daily living reveals changing mobility after surgery. Physiol Meas. Apr 3, 2019;40(3):035004. [doi: 10.1088/1361-6579/ab0d3e] [Medline: 30840937]
25. Ramamurthy SR, Chatterjee S, Galik E, et al. CogAx: early assessment of cognitive and functional impairment from accelerometry. In: 2022 IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE; 2022. [doi: 10.1109/PerCom53586.2022.9762401]
26. Garcia-Moreno FM, Bermudez-Edo M, Rodríguez-García E, Pérez-Mármol JM, Garrido JL, Rodríguez-Fórtiz MJ. A machine learning approach for semi-automatic assessment of IADL dependence in older adults with wearable sensors. Int J Med Inform. Jan 2022;157:104625. [doi: 10.1016/j.ijmedinf.2021.104625] [Medline: 34763192]
27. Pannurat N, Thiemjarus S, Nantajeewarawat E, Anantavrasilp I. Analysis of optimal sensor positions for activity classification and application on a different data collection scenario. Sensors (Basel). Apr 5, 2017;17(4):774. [doi: 10.3390/s17040774] [Medline: 28379208]
28. Mendes S, Queiroz J, Leitao P. Data driven multi-agent m-health system to characterize the daily activities of elderly people. In: 2017 12th Iberian Conference on Information Systems and Technologies (CISTI). IEEE; 2017. [doi: 10.23919/CISTI.2017.7975725]
29. Ibrahim A, Zhou Y, Jenkins ME, Trejos AL, Naish MD. Real-time voluntary motion prediction and Parkinson's tremor reduction using deep neural networks. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1413-1423. [doi: 10.1109/TNSRE.2021.3097007] [Medline: 34255631]
30. Maniar N, Chan SWT, Zulfikar W, Ren S, Xu C, Maes P. MemPal: leveraging multimodal AI and LLMs for voice-activated object retrieval in homes of older adults. In: Li T, Paternò F, Väänänen K, Leiva L, Spano D, Verbert K, editors. IUI '25: Proceedings of the 30th International Conference on Intelligent User Interfaces. Association for Computing Machinery; 2025:993-1015. [doi: 10.1145/3708359.3712151]
31. Nait Aicha A, Englebienne G, van Schooten KS, Pijnappels M, Kröse B. Deep learning to predict falls in older adults based on daily-life trunk accelerometry. Sensors (Basel). May 22, 2018;18(5):1654. [doi: 10.3390/s18051654] [Medline: 29786659]
32. Awais M, Palmerini L, Bourke AK, Ihlen EAF, Helbostad JL, Chiari L. Performance evaluation of state of the art systems for physical activity classification of older subjects using inertial sensors in a real life scenario: a benchmark study. Sensors (Basel). Dec 11, 2016;16(12):2105. [doi: 10.3390/s16122105] [Medline: 27973434]
33. Brodie MA, Wang K, Delbaere K, et al. New methods to monitor stair ascents using a wearable pendant device reveal how behavior, fear, and frailty influence falls in octogenarians. IEEE Trans Biomed Eng. Nov 2015;62(11):2595-2601. [doi: 10.1109/TBME.2015.2464689] [Medline: 26258936]
34. Hu Y, Bishnoi A, Kaur R, Sowers R, Hernandez ME. Exploration of machine learning to identify community dwelling older adults with balance dysfunction using short duration accelerometer data. Annu Int Conf IEEE Eng Med Biol Soc. Jul 2020;2020:812-815. [doi: 10.1109/EMBC44109.2020.9175871] [Medline: 33018109]
35. Mekruksavanich S, Phaphan W, Jitpattanakul A. Harnessing deep learning for activity recognition in seniors' daily routines with wearable sensors. In: 2024 47th International Conference on Telecommunications and Signal Processing (TSP). IEEE; 2024. [doi: 10.1109/TSP63128.2024.10605946]
36. Minici D, Cola G, Perfetti G, Espinoza Tofalos S, Di Bari M, Avvenuti M. Automated, ecologic assessment of frailty using a wrist-worn device. Pervasive Mob Comput. Oct 1, 2023;95(C):101833. [doi: 10.1016/j.pmcj.2023.101833]
37. Qiu H, Rehman RZU, Yu X, Xiong S. Application of wearable inertial sensors and a new test battery for distinguishing retrospective fallers from non-fallers among community-dwelling older people. Sci Rep. Nov 5, 2018;8(1):16349. [doi: 10.1038/s41598-018-34671-6] [Medline: 30397282]
38. Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, Freedson PS. Performance of activity classification algorithms in free-living older adults. Med Sci Sports Exerc. May 2016;48(5):941-950. [doi: 10.1249/MSS.0000000000000844] [Medline: 26673129]
39. Shang M, Dedeyne L, Dupont J, et al. Otago exercises monitoring for older adults by a single IMU and hierarchical machine learning models. IEEE Trans Neural Syst Rehabil Eng. 2024;32:462-471. [doi: 10.1109/TNSRE.2024.3355299] [Medline: 38231806]
40. Thralls KJ, Godbole S, Manini TM, Johnson E, Natarajan L, Kerr J. A comparison of accelerometry analysis methods for physical activity in older adult women and associations with health outcomes over time. J Sports Sci. Oct 2019;37(20):2309-2317. [doi: 10.1080/02640414.2019.1631080] [Medline: 31195893]
41. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of Wrist-Worn photoplethysmography devices to measure heart rate: A systematic review and meta-analysis. J Sports Sci. Sep 2020;38(17):2021-2034. [doi: 10.1080/02640414.2020.1767348] [Medline: 32552580]

42. Seo K, Hwang J, Kim K, Argrawal H, Yun JH, Yoon Kim H. End-to-end classification of cognitive impairment using daily-life gait data. In: 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2025. [doi: 10.1109/EMBC58623.2025.11254035]
43. Lim WS, Fan SP, Chiu SI, et al. Smartphone-derived multidomain features including voice, finger-tapping movement and gait aid early identification of Parkinson's disease. NPJ Parkinsons Dis. May 5, 2025;11(1):111. [doi: 10.1038/s41531-025-00953-w] [Medline: 40325040]
44. Guo WH, Yang XD, Ruan Z, et al. Early detection of Parkinson's disease: machine learning-based prediction of UPDRS Part III scores in de novo patients using smartphone assessments. J Parkinsons Dis. Sep 2025;15(6):1099-1110. [doi: 10.1177/1877718X251359494] [Medline: 40725991]
45. Zhu Z, Wu E, Leng P, Sun J, Ma M, Pan Z. Finger drawing on smartphone screens enables early Parkinson's disease detection through hybrid 1D-CNN and BiGRU deep learning architecture. PLoS One. 2025;20(7):e0327733. [doi: 10.1371/journal.pone.0327733] [Medline: 40658696]
46. Devito F, Gattulli V, Impedovo D. Care-MOVE: a smartphone-based application for continuous monitoring of mobility, environmental exposure and cognitive status in older patients. Appl Sci. 2026;16(3):1549. [doi: 10.3390/app16031549]
47. Shen Y, Park JH, Huang M, et al. Deep learning-based detection of cognitive impairment from passive smartphone sensing with routine-aware augmentation and demographic personalization. arXiv. Preprint posted online on Sep 27, 2025. [doi: 10.1109/BHI67747.2025.11269523]
48. Baum F, Musolino C, Gesesew HA, Popay J. New perspective on why women live longer than men: an exploration of power, gender, social determinants, and capitals. Int J Environ Res Public Health. Jan 14, 2021;18(2):661. [doi: 10.3390/ijerph18020661] [Medline: 33466763]
49. Nuzzo JL, Deaner RO. Men and women differ in their interest and willingness to participate in exercise and sports science research. Scand J Med Sci Sports. Sep 2023;33(9):1850-1865. [doi: 10.1111/sms.14404] [Medline: 37218686]
50. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. Apr 16, 2024;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
51. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth. Jun 3, 2021;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
52. Darley A, Carroll Á. Conducting co-design with older people in a digital setting: Methodological reflections and recommendations. Int J Integr Care. 2022;22(4):18. [doi: 10.5334/ijic.6546] [Medline: 36561633]
53. Bertolazzi A, Quaglia V, Bongelli R. Barriers and facilitators to health technology adoption by older adults with chronic diseases: an integrative systematic review. BMC Public Health. Feb 16, 2024;24(1):506. [doi: 10.1186/s12889-024-18036-5] [Medline: 38365698]
54. Mol A, Maier AB, van Wezel RJA, Meskers CGM. Multimodal monitoring of cardiovascular responses to postural changes. Front Physiol. 2020;11:168. [doi: 10.3389/fphys.2020.00168] [Medline: 32194438]
55. Tatipamula S. Real-time vs. batch data processing: when speed matters. World J Adv Res Rev. 2025;26(1):1612-1631. [doi: 10.30574/wjarr.2025.26.1.1213]
56. Regulation (EU) 2016/679 of the European parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (text with EEA relevance). European Union. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [Accessed 2026-11-02]
57. Shi Z, Du X, Li J, Hou R, Sun J, Marohabutr T. Factors influencing digital health literacy among older adults: a scoping review. Front Public Health. 2024;12:1447747. [doi: 10.3389/fpubh.2024.1447747] [Medline: 39555039]
58. Kim S, Park C, Park S, et al. Measuring digital health literacy in older adults: development and validation study. J Med Internet Res. Feb 5, 2025;27:e65492. [doi: 10.2196/65492] [Medline: 39908081]
59. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (text with EEA relevance.). European Union. 2017. URL: https://eur-lex.europa.eu/eli/reg/2017/745/oj [Accessed 2026-11-02]

Abbreviations

ACM: Association for Computing Machinery
ADL: activities of daily living
AUROC: area under the receiver operating characteristic curve
BP: blood pressure
CGWT: consumer-grade wearable technologies
GDPR: General Data Protection Regulation

HR: heart rate
IEEE: Institute of Electrical and Electronics Engineers
IMU: inertial measurement unit
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
PROBAST-AI: Prediction Model Risk-of-Bias Assessment Tool-AI
PROSPERO: International Prospective Register of Systematic Reviews
SpO_2: oxygen saturation
SR: systematic review
SWiM: synthesis without meta-analysis
TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis-Artificial Intelligence

Edited by Ivan Steenstra; peer-reviewed by Tiago Fonseca, Wenhao Qi; submitted 16.Feb.2026; initial comments to author 08.Apr.2026; final revised version received 18.Jun.2026; accepted 21.Aug.2026; published 26.Nov.2026

Please cite as:
Pera Calvi I, Andrade Bernardi F, Ferreira F, Lima VC, Guerche IC, Silva T, Dias DAM, da Silva TV, Rocha MCR, Almeida R, Fernández-Delgado M, Freitas A
Wearables and Machine Learning Use for Activity of Daily Life and Fall Management in Older Adults: Systematic Review
JMIR AI 2026;5:e93640
URL: https://ai.jmir.org/2026/1/e93640
doi: 10.2196/93640

© Izabela Pera Calvi, Filipe Andrade Bernardi, Francisco Ferreira, Vinícius Costa Lima, Isabela Carvalho Guerche, Tiago Silva, Duarte Azevedo Moura Dias, Teresa Valverde da Silva, Maria Cunha Ribeiro Rocha, Rute Almeida, Manuel Fernández-Delgado, Alberto Freitas. Originally published in JMIR AI (https://ai.jmir.org), 26.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on https://www.ai.jmir.org/, as well as this copyright and license information must be included.