Global Health Dynamics
https://ghd.cultechpub.com/ghd
E-ISSN: 3093-8066
Vol. 2, No. 2, November 2026
DOI: 10.64229/x1e6j250

Review

Hidden Frailty: Artificial Intelligence Detection of Early Functional Decline Before Clinical Frailty Becomes Apparent

Anyebe Daniel Ameh[1,*], Uchenna Ugochukwu Ozoekwe[2], IyinOluwa Ayoola Opadere[3]

[1]College of Professional Studies, University of Denver, Denver, Colorado, United States
[2]Department of Medical Rehabilitation, University of Nigeria Nsukka, Nigeria
[3]Department of Physiotherapy, State Hospital Ijaye, Abeokuta, Nigeria
[*]Corresponding author: Anyebe Daniel Ameh, ameh.daniel.hi@gmail.com

Abstract

Frailty is a state of reduced physiological reserve that increases the risk of falls, hospitalization, disability, and death in older adults. Established instruments, including the Fried phenotype, the Frailty Index, and the Clinical Frailty Scale, rely on episodic assessment against population thresholds, so decline is usually recognized only after measurable impairment has accumulated. In this narrative review, informed by structured literature searches, we propose hidden frailty as a methodological detection concept rather than a new clinical stage: a sustained, multidomain deviation from an individual's own functional baseline, detectable through continuous monitoring, which may begin while a person is still classified as robust or prefrail. We synthesize evidence on wearable, smartphone, ambient, and electronic-health-record signals of gait, activity, sleep, physiology, and speech, and appraise machine-learning models for frailty classification, prediction, and trajectory detection. Reported discrimination is highest for cross-sectional classification, where overlap between predictors and the frailty definition inflates performance, and lower for prospective prediction. Only one model has addressed trajectory-based detection (internal area under the receiver operating characteristic curve [AUC] 0.70, no external validation), and calibration, decision-curve analysis, and economic evaluation are rarely reported. We propose a six-stage framework that specifies candidate variables, provisional decision criteria, clinical responses, and research priorities for each stage, and outline a pragmatic trial to test it. Because the premise currently rests largely on extrapolation from cross-sectional and short-term data, hidden frailty remains a testable hypothesis whose value depends on whether earlier detection, delivered equitably and with respect for privacy, improves functional outcomes.

Keywords

Frailty, Artificial intelligence, Machine learning, Functional decline, Digital biomarkers, Wearable sensors, Older adults, Digital phenotyping

Article History

Received: 26 September 2026 Revised: 03 November 2026

Accepted: 09 November 2026 Available Online: 16 November 2026

Copyright

© 2026 by the authors. This article is published by the Cultech Publishing Sdn. Bhd. under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

Frailty is increasingly understood not as a fixed diagnostic label but as a dynamic, multidimensional state of diminished physiological reserve and heightened vulnerability to adverse outcomes following even minor stressors, such as an infection, a medication change, a fall, or a brief hospitalization. Population aging is accelerating worldwide, and with it the number of older adults living with, or at risk of, frailty [1]. A pooled analysis of 240 studies spanning 62 countries and 1.76 million participants estimated frailty prevalence at 12% using physical-phenotype criteria but 24% using a deficit-accumulation index, a twofold difference driven entirely by which instrument was applied to comparable populations [2], underscoring from the outset how much definitional choice shapes what "frailty" is taken to mean. This vulnerability arises from the cumulative, interacting decline of multiple physiological systems, including the musculoskeletal, cardiovascular, metabolic, immune, and neurocognitive domains, such that the capacity to maintain homeostasis under stress is progressively eroded.

Conventional frailty assessment tools remain indispensable to geriatric practice and continue to anchor clinical decision-making and research. The Fried frailty phenotype operationalizes frailty through five physical criteria, namely unintentional weight loss, exhaustion, weakness, slow gait speed, and low physical activity, classifying individuals as robust, prefrail, or frail [3]. The Frailty Index, grounded in a deficit-accumulation model, aggregates dozens of age-related health deficits into a continuous measure of biological vulnerability [4,5]. The Clinical Frailty Scale offers a pragmatic, judgment-based ordinal classification widely used in acute and perioperative settings [6]. Each instrument has demonstrated predictive validity for adverse outcomes and has meaningfully advanced geriatric risk stratification.

Yet all three share a fundamental limitation: they are episodic. Frailty status is typically ascertained at a single clinical encounter, often months or years apart, and classification itself presupposes that measurable functional impairment has already accumulated to a detectable threshold. By the time an older adult is reclassified from prefrail to frail, the underlying physiological erosion may have been progressing silently for a considerable period, and clinicians are frequently left responding to frailty only once it has become clinically manifest.

We therefore propose the concept of hidden frailty to describe this preclinical interval, a state in which subtle, measurable deterioration in functional reserve is already underway but has not yet crossed the thresholds recognized by conventional instruments. Characterizing this state requires moving beyond periodic clinical assessment toward continuous, ecologically valid measurement of everyday function. Advances in wearable technologies, smartphones, ambient sensors, electronic health records, computer vision, and machine learning now make it feasible to capture gait, activity, sleep, physiological, and behavioral signals unobtrusively and repeatedly over time. Importantly, hidden frailty is not proposed as an additional biological stage interposed between robustness and prefrailty. It is a methodological concept describing what becomes observable when an individual's function is measured continuously and compared with that individual's own baseline, rather than intermittently against population cut-points; as discussed in Section 2.4, the window it describes may overlap with states currently classified as robust or prefrail.

This narrative review examines the potential of artificial intelligence and digital biomarkers to identify early functional decline before clinically apparent frailty, critically evaluates limitations of the current evidence, and proposes a trajectory-based framework for future frailty detection, situating the discussion within the broader technical, regulatory, and health-systems landscape in which such tools would ultimately need to operate.

One finding frames the remainder of this review and is therefore stated at the outset: in the most recent systematic review of machine-learning frailty models, only a single model attempted a trajectory-based task, classifying rapid versus stable growth in the Frailty Index, and it reported an internal area under the receiver operating characteristic curve (AUC) of 0.70 without any external validation [7]. Direct evidence that artificial intelligence can detect hidden frailty is therefore currently very limited, and much of what follows draws on adjacent evidence from cross-sectional classification and prospective prediction, which must be interpreted with this distinction in mind.

Scope and Methods of this Review: This article is a narrative (non-systematic) review, prepared with reference to the SANRA criteria for the quality of narrative reviews [8]. To reduce the risk of arbitrary study selection, the literature base was identified through structured searches of PubMed/MEDLINE, Scopus, Web of Science, and Dimensions, conducted between January and August 2026 for English-language articles published from 1999 to 2026, using combinations of terms for frailty (frailty, prefrailty, frailty index, functional decline, intrinsic capacity), digital measurement (wearable, accelerometer, smartphone, ambient sensor, computer vision, electronic health record, digital biomarker, digital phenotyping), and analytic methods (machine learning, artificial intelligence, deep learning, prediction model). Reference lists of the included systematic and scoping reviews were also hand-searched, and methodological and reporting-standard references were cited as needed irrespective of publication date.

Studies were prioritized for detailed synthesis, and for inclusion in Table 1, if they met at least one of three criteria: (i) they defined a frailty construct used as a reference standard in the digital literature; (ii) they reported quantitative discrimination or association statistics for a digital or physiological signal against a frailty-related outcome in older adults; or (iii) they were systematic, scoping, or meta-analytic syntheses of frailty prediction models or digital frailty measurement. Where a recent systematic synthesis existed, it was preferred over individual primary studies, and findings that ran counter to the central thesis of this review, such as non-significant balance measures, moderate correlations for

continuous severity, and the superior median performance of regression over machine-learning models, were retained rather than excluded. A formal duplicate risk-of-bias assessment of every primary study was not undertaken; instead, we report the formal PROBAST (Prediction model Risk Of Bias ASsessment Tool) [9] appraisals already conducted by the systematic reviews we cite [7], and we provide a structured, PROBAST-informed appraisal of the key methodological features of each study in Table 1, presented in Table 2. The implications of the narrative design for selection and publication bias are discussed in Sections 7.8 and 7.9.

Table 1. Selected quantitative evidence underlying the hidden frailty synthesis.

Study	Population (N)	Domain	Method	Key quantitative finding	External validation
Fried et al. 2001 [3]	5,317 (CHS)	Definitional (phenotype)	5-criterion phenotype	6.9% prevalence; HR 1.3-4.5 for falls, disability, hospitalization, death	N/A - criterion-defining study
Rockwood & Mitnitski 2007 [4]	33,581 + 2,573 (7 cohorts)	Definitional (deficit accumulation)	Frailty Index	Deficit-accumulation slope 0.029/year; characteristic submaximal ceiling	Multiple pooled cohorts
O'Caoimh et al. 2021 [2]	1,755,497 (240 studies, 62 countries)	Cross-definition prevalence	Meta-analysis	12% (phenotype) vs. 24% (Frailty Index) prevalence - same populations, different instrument	N/A - meta-analysis
Studenski et al. 2011 [10]	34,485 (9 cohorts)	Gait speed	Pooled survival analysis	HR 0.88 per 0.1 m/s gait speed increment	Multi-cohort pooled
Schwenk et al. 2015 [11]	125	Gait, balance, activity	In-home wearable sensors	AUC 0.802 (gait speed) - 0.857 (stride length); balance sway not significant	Single cohort, none
Mou et al. 2026 [12]	Meta-analysis, multiple cohorts	Falls prediction	Wearable devices	Pooled AUC ≈ 0.85	Meta-analytic
Culos et al. 2026 [13]	437 (TSHA)	Physical activity	XGBoost	Spearman's $\rho = 0.55$ vs. continuous frailty score	None reported
Varadhan et al. 2009 [14]	276 (WHAS-I)	Heart-rate variability	PCA-derived composite	$\beta = -0.73$ (frailty); $\beta = -0.60$ (5-yr mortality)	Single cohort
Da Cunha et al. 2026 [15]	271	Speech	Multimodal ML	Mean AUC = 0.91 across 10 physical domains	Single cohort, none
Kim et al. 2026 [7]	14 studies (1,230-86,133 pooled)	ML classification/prediction (synthesis)	Systematic review	AUC 0.70-0.98 (classification); 0.58-0.85 (prediction)	Only 2/14 low risk of bias + low applicability (PROBAST)
Xiao et al. 2026 [16]	31 models, 30 studies	Frailty-onset prediction (synthesis)	Scoping review	Median AUC 0.860 (regression) vs. 0.813 (machine learning)	22.6% externally validated

Note: AUC, area under the receiver operating characteristic curve; CHS, Cardiovascular Health Study; HR, hazard ratio; ML, machine learning; PROBAST, Prediction model Risk Of Bias ASsessment Tool; TSHA, Toledo Study for Healthy Aging; WHAS-I, Women's Health and Aging Study I.

Table 1 summarizes the population sizes, methods, and key quantitative findings of the studies central to the synthesis developed in Sections 3, 4, and 7, including the discrepancies and contradictions those sections discuss in more depth. Table 2 summarizes the design, validation status, risk of feature-outcome overlap, and principal limitation of the same studies.

Table 2. Structured, PROBAST-informed appraisal of the studies summarized in Table 1.

Study	Design	Validation	Feature-outcome overlap (circularity risk)	Main limitation for hidden-frailty inference	Overall concern
Fried et al. 2001 [3]	Prospective cohort	Criterion-defining	Not applicable	Defines between-person categories; not designed to detect within-person change	Low (as reference standard)
Rockwood & Mitnitski 2007 [4]	Multiple cohorts	Multiple cohorts	Not applicable	Continuous, but measured episodically	Low (as reference standard)
O'Caoimh et al. 2021 [2]	Meta-analysis of population studies	Not applicable	Not applicable	Marked heterogeneity of frailty definitions	Low to moderate
Studenski et al. 2011 [10]	Pooled prospective cohorts	Multi-cohort	Low (gait speed vs. survival)	Clinic-measured gait speed, not wearable-derived	Low
Schwenk et al. 2015 [11]	Cross-sectional (baseline)	None; single cohort (n = 125)	High: gait speed is a phenotype criterion	Small sample; no longitudinal outcome	High
Mou et al. 2026 [12]	Meta-analysis	Pooled	Low to unclear	Outcome is falls, not frailty; between-study heterogeneity	Moderate
Culos et al. 2026 [13]	Cross-sectional model development; internal subsampling	No external validation	Unclear: activity may overlap with frailty-score components	Moderate correlation only; single cohort	Moderate to high
Varadhan et al. 2009 [14]	Cross-sectional, with prospective mortality	Single cohort	Low	Women only; modest sample	Moderate
Da Cunha et al. 2026 [15]	Cross-sectional	No external validation	Low (speech vs. physical function)	Single cohort; outcomes are physical-function domains rather than frailty	High (unreplicated)
Kim et al. 2026 [7]	Systematic review with formal PROBAST	Not applicable	High for the top-performing classification models (AUC 0.97-0.98)	Only 2 of 14 studies at low risk of bias with low applicability concern	Synthesis (formal appraisal)
Xiao et al. 2026 [16]	Scoping review without formal risk-of-bias assessment	Not applicable	Not assessed	22.6% externally validated; calibration reported for 74.2% of models (50% of ML models); decision-curve analysis for 32.3%	Synthesis (no formal appraisal)

Note: Judgements are the authors' structured narrative appraisal, informed by the four PROBAST domains (participants, predictors, outcome, analysis); they are not a substitute for a formal, duplicate PROBAST assessment. AUC, area under the receiver operating characteristic curve; ML, machine learning; PROBAST, Prediction model Risk Of Bias Assessment Tool.

2. Frailty as a Dynamic Process: The Case for Earlier Detection

2.1 Frailty Beyond Binary Classification

Much of clinical practice treats frailty as a binary condition: a patient either meets criteria for frailty or does not. This framing, while operationally convenient, obscures the underlying biology. Frailty more accurately represents a point along a continuum that extends from robust functioning, through subtle and often imperceptible decline, into prefrailty, established frailty, and eventually severe vulnerability [17]. Movement along this continuum is rarely abrupt; it typically unfolds gradually, punctuated by acute stressors that reveal reserve already lost rather than reserve newly depleted. Treating frailty as frail-versus-non-frail forces an artificial discontinuity onto what is, physiologically, a gradual erosion of adaptive capacity, and in doing so may delay recognition of the earliest, most modifiable stages of decline.

2.2 Functional Reserve and Resilience

Functional reserve, the physiological buffer that allows recovery from stressors, declines progressively across mobility, physical capacity, cognitive resilience, and homeostatic recovery following illness or injury. This decline may begin well before it manifests as the overt signs captured by conventional instruments: a person may still walk at a normal pace and pass standard screening while experiencing measurable reductions in gait efficiency, activity tolerance, or recovery speed, reductions frequently below the resolution of episodic examination yet potentially predictive of subsequent transition to clinically apparent frailty. This is complicated further for older adults simultaneously managing multimorbidity, since chronic-disease accumulation and frailty appear to interact and reinforce one another rather than operating independently [18].

2.3 The Clinical Assessment Blind Spot

The traditional model of frailty assessment follows a straightforward sequence: periodic assessment, threshold reached, frailty identified. This model depends entirely on the timing of clinical contact and is structurally incapable of detecting deterioration between encounters. A complementary digital model becomes possible with continuous monitoring: trajectory change, early signal, clinical reassessment. The object of interest shifts from a single cross-sectional measurement to the shape of change over time, and the clinical question shifts from how a person compares with others to how they compare with themselves previously. This distinction underlies the central argument of this review: artificial intelligence may ultimately prove more valuable for identifying meaningful change in an individual's functional trajectory than for assigning static frailty labels at a single point in time. This is, however, a hypothesis rather than a demonstrated advantage: as Section 4.2 shows, very few existing studies have yet operationalized within-person change, and the case for artificial intelligence in this role currently rests on its suitability for high-dimensional, longitudinal data rather than on direct evidence of superior early detection.

2.4 Positioning Hidden Frailty Relative to Existing Frailty Constructs

Hidden frailty overlaps conceptually with several established constructs, and its distinctiveness lies primarily in how decline is measured rather than in a separate biological stage (Table 3). Under the Fried phenotype, prefrailty is defined by the presence of one or two of the five criteria [3]; it is a between-person category assigned at a single assessment, and individuals move between robust, prefrail, and frail states in both directions over time [19]. The Frailty Index is continuous and can register small increments of deficit accumulation below any categorical cut-point [4,5], but in practice it is measured episodically, often years apart, and its deficits are largely clinical or self-reported, so a change of one or two deficits between assessments is difficult to distinguish from measurement noise. The World Health Organization construct of intrinsic capacity similarly emphasizes declining capacities across locomotor, vitality, cognitive, psychological, and sensory domains before dependency emerges [20], and "early functional decline" is used descriptively across geriatric research without a standard operational definition.

Table 3. Hidden frailty in relation to existing frailty-related constructs.

Construct	Reference point	Measurement frequency	Nature	What the hidden frailty approach adds
Prefrailty (Fried phenotype) [3]	Population cut-points; 1-2 of 5 criteria present	Episodic (clinical or research visit)	Categorical clinical state	Detects within-person deterioration in criterion domains (e.g., gait, activity) before a cut-point is crossed, or progression within prefrailty
Frailty Index [4,5]	Proportion of deficits present (continuous)	Episodic; typically years apart	Continuous risk measure	Higher temporal resolution; performance-based signals; within-person reference
Intrinsic capacity (WHO) [20]	Capacity across five domains	Episodic screening	Functional-capacity construct	Continuous, passive measurement of locomotor and vitality-related domains
Early functional decline (descriptive)	Variable; often self-report or performance tests	Variable	Descriptive term without a standard definition	An operational definition: sustained, multidomain deviation from a personal baseline
Hidden frailty (proposed)	The individual's own digital baseline	Continuous or high-frequency	Methodological detection concept (hypothesis), not a diagnostic stage	Requires clinical confirmation (Section 5.4) and prospective validation against established measures and outcomes

Note: WHO, World Health Organization.

We therefore position hidden frailty explicitly as a methodological and operational concept for detecting earlier change within the established frailty progression process, not as a new clinical state or diagnostic stage. Three features distinguish it from existing constructs. First, the reference point is the individual's own baseline rather than a population cut-point. Second, measurement is continuous or high-frequency rather than episodic, so the object of inference is the slope and persistence of change rather than a level at a single time. Third, deviation must be multidomain and sustained, as set out in Sections 5.3 to 5.5. Consequently, hidden frailty does not occupy a fixed position on the robust-prefrail-frail sequence: a person meeting no Fried criteria, or one criterion and therefore already classified as prefrail, could equally show a hidden frailty signal if their trajectory is deteriorating relative to their own baseline. Figure 1A accordingly depicts hidden frailty as a within-person detection window overlapping the robust and early prefrail states, rather than as a discrete stage preceding prefrailty. The added value of the concept is therefore testable: it should be judged by whether within-person digital signals anticipate transitions on established measures, such as an increase in the number of Fried criteria or in the Frailty Index, earlier or more accurately than repeated conventional assessment alone.

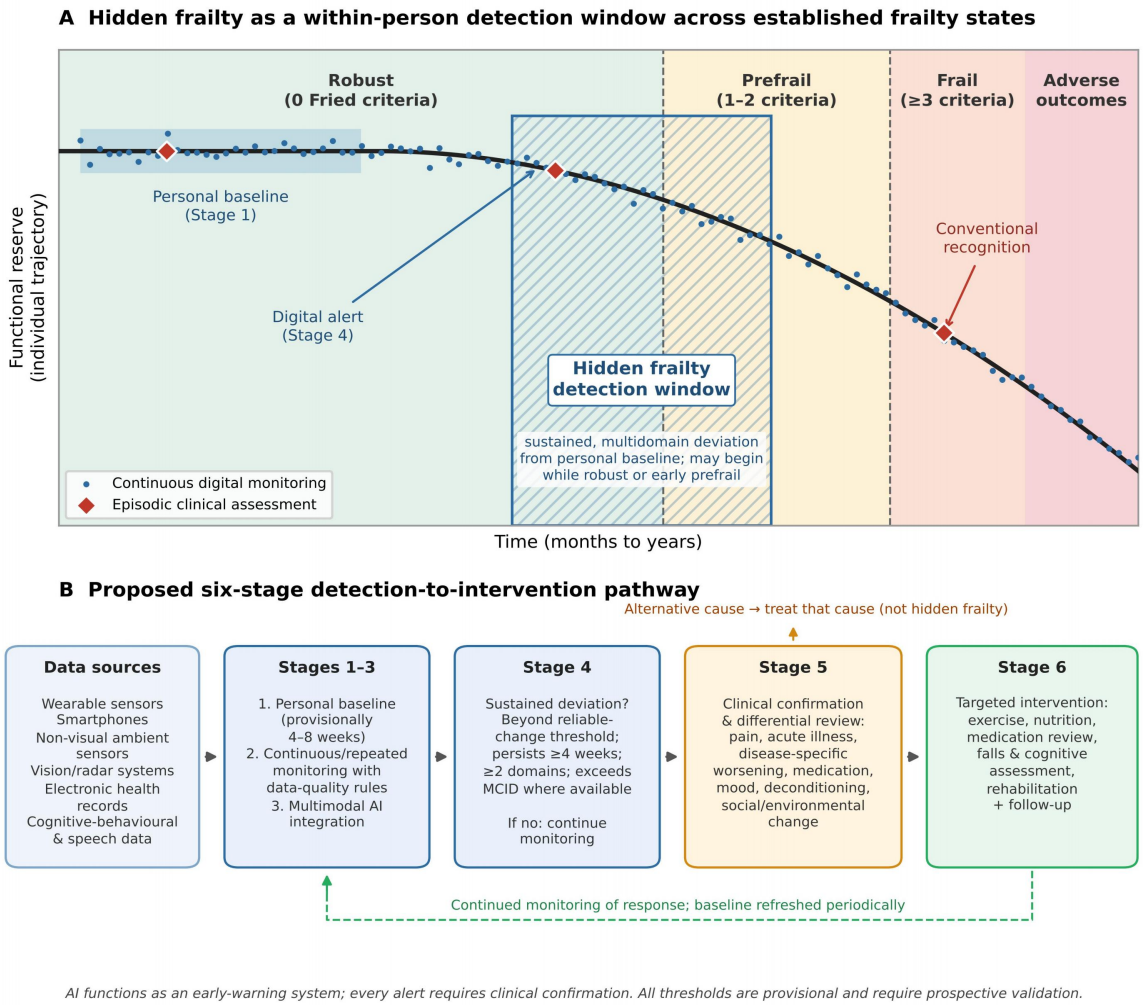


Figure 1. Hidden frailty as a within-person detection window, and the proposed six-stage detection-to-intervention pathway. (A) An individual's functional-reserve trajectory (black line) is shown against established frailty states defined by conventional population thresholds (dashed lines; Fried phenotype criteria shown for illustration). Continuous digital monitoring (blue dots) allows a personal baseline to be established and a sustained, multidomain deviation from that baseline to be detected (hidden frailty detection window), which may begin while the person is still classified as robust or early prefrail; episodic clinical assessments (red diamonds) recognize decline only once conventional thresholds are crossed. Hidden frailty is therefore depicted as a methodological detection window that overlaps existing states, not as a discrete stage between robustness and prefrailty. (B) The six stages of the proposed framework, with provisional decision criteria; detected deviations require clinical confirmation and differential review before being attributed to reduced physiological reserve. All thresholds shown are provisional and require prospective validation. MCID, minimal clinically important difference.

3. Digital Signatures of Early Functional Decline

If hidden frailty exists as a measurable phenomenon, it should leave detectable traces in everyday function well before it becomes clinically apparent. This section synthesizes the categories of digital signals, spanning wearable, smartphone, and ambient sources, that artificial intelligence systems could plausibly use to detect such early decline.

3.1 Mobility, Gait, and Falls

Gait integrates musculoskeletal strength, balance, cardiorespiratory capacity, and neuromotor control into a single observable behavior. In a pooled analysis of nine cohorts totaling 34,485 community-dwelling adults followed for a mean of 12.2 years, each 0.1 m/s increment in gait speed was associated with a mortality hazard ratio of 0.88, and predicted 10-year survival at age 75 spanned from 19% to 87% in men and 35% to 91% in women, a model whose accuracy matched considerably more complex risk scores incorporating comorbidity and self-reported function [10]. Whether wearable-derived gait metrics can substitute for this clinic-based measurement, and detect deterioration earlier, is the more specific question digital biomarker research must answer. In the 125-person Arizona Frailty Cohort Study, in-home wearable sensors discriminated prefrail from nonfrail participants using gait speed alone with an AUC of 0.802, and stride length discriminated among all three frailty groups with an AUC of 0.857; balance-derived sway measures, however, showed no significant discrimination at all, suggesting not every mobility domain carries equal early-warning value, a distinction narrative treatments of "gait and balance" as one category tend to obscure [11]. Because gait speed is itself one of the five phenotype criteria used to define the reference standard, part of this discrimination may reflect feature-outcome overlap

rather than independent early-warning value (Table 2). Wearable-derived features also support prospective fall prediction with reported discrimination around an AUC of 0.85 [12], while camera- and radar-based ambient systems offer a complementary, device-free route to detecting falls indoors, avoiding the adherence problems wearables face [21].

3.2 Physical Activity and Daily Behavior

Passively collected activity data (e.g., step counts, sedentary time, activity fragmentation, and time away from home) are appealing because they require no active gait test. Yet direct evidence of their predictive value is more modest than the gait findings above: in a model trained on 437 participants of the Toledo Study for Healthy Aging using wrist accelerometry and gradient-boosted trees, activity features correlated with a continuous frailty score at only a moderate level, Spearman's ρ of 0.55, evaluated solely by repeated subsampling within that single cohort, without external validation [13]. This gap between a moderate within-cohort correlation and the considerably higher classification statistics reported elsewhere in this literature (Section 4) is itself informative: predicting graded frailty severity from activity data appears harder than binary classification, and claims about activity-based biomarkers should be read with that distinction in mind. As with gait, the more defensible signal is within-person change rather than any population-referenced threshold. Reduced activity is also the signal most open to explanations unrelated to frailty, such as pain, weather, bereavement, relocation, or personal preference, a problem addressed in Section 5.4.

3.3 Sleep and Circadian Patterns

Sleep architecture and circadian regulation frequently deteriorate alongside declining functional reserve, manifesting as fragmentation, altered duration, irregular timing, and increased nocturnal activity. In the Osteoporotic Fractures in Men cohort, actigraphy- and polysomnography-measured sleep disturbance was independently associated with prevalent and incident frailty among community-dwelling older men, even after adjustment for comorbidity [22]. These signals are readily captured passively but are highly nonspecific, since sleep disruption accompanies pain, mood disorders, and environmental disturbance unrelated to frailty, and are best treated as contributory rather than independently diagnostic.

3.4 Physiological Signals

In 276 women aged 65 to 101 in the Women's Health and Aging Study, a composite heart-rate variability measure derived from principal-components analysis of ambulatory electrocardiograms was more strongly associated with frailty status ($\beta = -0.73$, $p < 10^{-5}$) than any individual traditional HRV index, and the same composite was the strongest available predictor of five-year mortality among the physiological measures examined ($\beta = -0.60$, $p < 10^{-6}$) [14]. This suggests physiological signals may carry genuine, non-redundant information about frailty risk, but also that single conventional indices likely underperform composite or learned representations, an argument for the multimodal integration discussed in Section 4.4 rather than for physiological monitoring in isolation.

3.5 Cognitive, Speech, and Behavioral Signals

A newer and still less validated category concerns speech, smartphone interaction, and social-engagement signals. In 271 community-dwelling older adults, acoustic, temporal, and linguistic features extracted from one minute of spontaneous speech predicted ten distinct physical function domains, including gait speed and handgrip strength, with a mean AUC of 0.91 across domains, and eliminated false negatives for two of the ten measures [15]. This is a considerably higher discrimination statistic than the mobility- or activity-based models above, and should be interpreted cautiously for the same reason flagged throughout this section: it comes from a single cohort, and the broader digital phenotyping paradigm from which it draws was itself developed and validated primarily in behavioral health rather than physical function [23], so independent replication in a second, demographically distinct sample would meaningfully strengthen the claim before it is treated as an established digital biomarker.

Taken together, these five domains illustrate a consistent pattern rather than five independent success stories: reported discrimination is highest for narrow, well-instrumented, single-cohort classification tasks (Schwenk's stride-length AUC of 0.857, Da Cunha's speech AUC of 0.91) and considerably more modest for the harder, arguably more clinically relevant task of tracking continuous severity or change over time (Culos's ρ of 0.55). Considered individually, none of these signatures is likely to be sufficient to identify hidden frailty reliably. Multimodal combinations, integrating mobility, activity, sleep, physiology, and cognitive-behavioral signals, are a more plausible route to the earlier, more gradual signal this review is concerned with, consistent with the inherently multidimensional nature of frailty itself.

4. Artificial Intelligence for Frailty Detection and Prediction

4.1 Current Machine-Learning Applications

A systematic review of fourteen studies applying machine-learning methods, logistic regression, random forests, gradient-boosted trees such as CatBoost, and neural architectures, to frailty classification found development cohorts ranging from 1,230 to 86,133 participants and discrimination for cross-sectional classification spanning an AUC of 0.70 to 0.98 [7]. That upper value is worth pausing on: the same review attributed its highest-performing models, CatBoost and logistic regression both reaching AUCs of 0.97-0.98, to likely overlap between input predictors and the frailty definition being predicted rather than genuine predictive insight, a concrete instance of the circular-prediction risk discussed further in Section 7.4 [7]. Models anticipating future, rather than current, frailty performed markedly less well: internal AUCs of 0.70-0.81, falling to 0.58-0.85 under same-cohort temporal validation [7]. A larger scoping review of thirty-one models published between 2018 and 2026 reported a further, counterintuitive pattern: regression-based models achieved a higher median AUC (0.860) than machine-learning models (0.813), with "no systematic performance advantage" observed for the more complex methods [16], complicating any assumption that algorithmic sophistication automatically improves prediction. Wearable-based multimodal models combining gait, activity, and physiological features exist but remain rare and, as Section 3.2 discussed, perform considerably less well in predicting continuous frailty severity than in binary classification [13].

4.2 Frailty Classification Versus Frailty Prediction

A central conceptual distinction separates three related but fundamentally different tasks, and the evidence above maps onto it directly. Current frailty classification asks whether a person is currently frail, the task responsible for the highest, most circularity-prone AUCs reported above. Future frailty prediction asks whether a person will develop frailty at some future time point, the task whose AUCs drop to 0.58-0.85. Early functional-decline detection asks whether a person's functional trajectory is beginning to deteriorate, independent of whether that deterioration culminates in a formal diagnosis; only a single model in the systematic review above attempted anything resembling this trajectory-based task, reporting an internal AUC of 0.70 with no external validation at all [7]. That model classified rapid versus stable growth in the Frailty Index using predominantly questionnaire- and assessment-derived predictors, including age, activities of daily living, cognitive function, chronic disease burden, and lifestyle factors [7]; its temporal resolution was therefore that of scheduled study assessments rather than continuous sensor data, and it addresses trajectory classification rather than the high-frequency, within-person detection envisaged here. Much of the existing literature conflates these tasks, evaluating models built for one purpose against outcomes appropriate to another. Disentangling them is essential, since the tools, data structures, and validation approaches appropriate to each differ substantially, and since, on the evidence above, performance degrades sharply and validation becomes sparser at precisely the point the literature moves from classification toward the trajectory-based question this review argues is most clinically valuable.

4.3 Longitudinal and Time-Series Modeling

Detecting early functional decline is fundamentally a longitudinal problem. Trajectory analysis, anomaly detection, change-point detection, recurrent architectures, and other temporal deep-learning methods characterize how a signal evolves within an individual, rather than comparing individuals cross-sectionally, and are in principle considerably better suited to identifying hidden frailty than conventional classifiers built for a single moment in time.

4.4 Multimodal and Federated AI

Frailty is a multidimensional syndrome, so its earliest manifestations are unlikely to be adequately captured by any single data stream; multimodal approaches integrating gait, activity, sleep, physiology, cognition, and clinical information offer a more coherent match to the underlying construct, though harmonizing differing sampling frequencies and missingness patterns across modalities remains demanding. Because no single institution is likely to hold sufficient longitudinal, multimodal data to train and externally validate such models robustly, federated learning, in which models are trained collaboratively across multiple sites without centralizing raw patient data, has been proposed as a practical route to the large, diverse training populations this problem requires while preserving local data governance [24].

4.5 Explainability and Clinical Trust

A further, often neglected requirement is interpretability. Of the fourteen studies in the systematic review discussed above, only two employed formal explainability methods such as SHAP or SAGE, with most others relying on simpler feature-importance rankings or providing minimal predictor-level transparency at all [7]. Clinicians are unlikely to act on an opaque model's output, particularly for a preliminary construct like hidden frailty that lacks an established clinical referent, and explainability has been identified as a multidisciplinary requirement spanning technical, clinical, ethical, and

regulatory considerations rather than a purely computational add-on [25]. A model that flags trajectory deviation without indicating which underlying domain drove the signal offers a clinician little guidance on what to assess next.

4.6 Calibration, Clinical Utility, and Performance Under Validation

Discrimination alone cannot establish clinical usefulness, and reporting of other performance dimensions is incomplete. In the scoping review of thirty-one frailty-onset models, 23 (74.2%) reported some measure of calibration, but only 6 of 12 (50%) machine-learning models did so, and decision-curve analysis was reported for only 10 models (32.3%) [16]. The systematic review of fourteen machine-learning studies did not tabulate calibration or decision-curve reporting, instead identifying both as priorities for future work [7]. A quantitative synthesis of calibration or net benefit across the cited models is therefore not currently possible, because the required statistics are unavailable for most models. Paired internal and external performance estimates are similarly scarce: internal AUCs of 0.70-0.81 for incident frailty prediction fell to 0.58-0.85 under same-cohort temporal validation [7], and only 22.6% of models in the scoping review had been externally validated at all [16]. Methodological work outside frailty shows that discrimination is typically lower when prediction models are validated in new populations, particularly by independent investigators [26], so the internal estimates reported in this literature are best regarded as upper bounds. Decision-curve analysis, which estimates net benefit across clinically plausible decision thresholds [27], is particularly relevant to hidden frailty, where most alerts would trigger a low-cost clinical review rather than a definitive intervention, and where the acceptable ratio of false to true alerts must therefore be made explicit.

5. From Population Thresholds to Individual Functional Trajectories

5.1 Limitations of Population-Level Thresholds

Virtually all conventional frailty instruments rely on population-derived thresholds: a gait speed below a fixed cut-point, a weight loss exceeding a fixed percentage, a deficit count above a fixed proportion. These thresholds are calibrated against population distributions and are therefore insensitive to an individual's own starting point. An older adult whose habitual walking speed is exceptionally high may experience clinically meaningful decline without ever crossing a population-defined impairment threshold. Conversely, someone whose baseline function was always modest may be flagged as impaired despite exhibiting no meaningful change at all. Population thresholds can be simultaneously insensitive to genuine deterioration in high-functioning individuals and non-specific for those whose baseline differs from the reference population.

5.2 Individualized Baselines

An alternative approach, made feasible by continuous digital monitoring, is to establish a personal functional baseline for each individual and to evaluate subsequent measurements against that baseline rather than against a fixed population cut-point, an extension to physical function of the broader digital phenotyping paradigm developed in behavioral health [23]. This suggests a plausible operational sequence: a personal baseline is established through an initial period of measurement; repeated measurement continues over time; statistical deviation from that baseline is detected; if deterioration proves persistent rather than transient, it prompts clinical review.

5.3 Detecting Meaningful Deviation

Individualized monitoring introduces its own analytic challenge: distinguishing genuine, sustained deterioration from the many forms of noise that affect any longitudinal signal. Normal day-to-day variation, transient decline during acute illness, environmental changes such as travel or relocation, changes in device wear, and temporary inactivity unrelated to health status can all mimic true functional decline. A credible detection system must differentiate persistent trajectory change from these transient sources, typically by requiring deviation to be sustained across a defined minimum duration and consistent across more than one domain before it is treated as clinically meaningful. Several established statistical approaches could formalize this requirement. A personal reference range can be estimated from the baseline period, and subsequent change expressed as a reliable change index, which scales the observed change by the measurement error of the individual's own data [28]; sequential methods such as cumulative-sum (CUSUM) or exponentially weighted moving-average control charts, and Bayesian change-point models, can then test whether a deviation is persistent rather than isolated. As a provisional and testable criterion, we suggest that a deviation be regarded as potentially meaningful only when (i) it exceeds the individual's reliable-change threshold, for example a reliable change index beyond ±1.96; (ii) it persists for a pre-specified period long enough to exclude typical recovery from minor acute illness, provisionally at least four weeks; (iii) it is present in at least two domains; and (iv) where an anchor exists, it exceeds an established minimal clinically important difference, such as approximately 0.05 m/s for a small and 0.10 m/s for a substantial change in gait speed [29]. These thresholds are proposals for empirical testing rather than validated criteria, and their sensitivity and false-alert rates must be estimated in prospective cohorts.

5.4 Interpreting Digital Decline: Reduced Reserve Versus Alternative Explanations

A reduction in daily activity or walking performance does not, by itself, indicate declining physiological reserve. Activity may fall because of pain, an acute illness, a change in living environment, weather or season, reduced social participation, caregiving demands, or personal preference, and functional decline may reflect worsening of a specific disease, such as heart failure, osteoarthritis, depression, or Parkinson's disease, or deconditioning after a period of inactivity, rather than frailty progression. Digital signals are therefore best treated as evidence that something has changed; attributing that change to reduced reserve is an inference that requires additional support.

Distinguishing physiological decline from transient, situational shifts is the most demanding analytic problem the framework faces, because daily behavioral data are inherently noisy and many confounders, such as weather, intercurrent pain, or temporary caregiving duties, act on the same timescale as early decline. Three complementary strategies could reduce, although not eliminate, this problem. First, known external drivers can be measured and modeled directly rather than treated as noise: local meteorological data and day length can be linked by date and location, and calendar effects such as weekends, holidays, and season can be included as covariates in each individual's trajectory model, so that deviation is assessed relative to the activity expected under prevailing conditions. Second, person-reported context, captured through brief, low-burden prompts (e.g., a weekly one- or two-item check-in on pain, illness, caregiving, or travel) or through linked electronic-health-record events, can be used to label periods of known situational disruption, to suppress alerts during them, and to test whether the trajectory subsequently returns to baseline. Third, the analysis can give greater weight to signals that are less dependent on circumstance: capacity-related measures derived from bouts of walking, such as gait speed and stride characteristics, are likely to be less affected by weather or daily obligations than total activity volume, and the requirement for concordant deviation across physiologically distinct domains sustained for at least four weeks (Section 5.3) is intended to filter out short-lived, single-domain shifts. Even with these safeguards, some transient changes will remain indistinguishable from early decline on digital data alone. For this reason, the framework treats every alert as a prompt for clinical confirmation rather than as a diagnosis (Section 6, Stage 5), and the proportion of alerts attributable to situational rather than physiological causes must be estimated empirically in prospective studies.

We propose that digitally detected decline be interpreted as consistent with reduced physiological reserve, or progression toward frailty, only when several conditions converge: (i) persistence beyond the expected recovery period of any identifiable acute stressor; (ii) concordance across physiologically distinct domains, for example gait together with sleep or heart-rate variability, rather than a fall in activity volume alone, since capacity-related measures such as gait speed and stride characteristics are less sensitive to preference than step counts; (iii) the absence, on clinical review, of a sufficient alternative explanation, including new pain, acute or exacerbated disease, medication change, depression, or environmental or social change, with active screening for psychological distress rather than reliance on recorded diagnoses alone, since functional impairment may signal distress in older adults who have no depression diagnosis [30]; (iv) impaired recovery dynamics, that is, slower or incomplete return to baseline after minor stressors, which has been proposed as a direct marker of reduced resilience [31]; and (v) corroboration by conventional performance measures, such as gait speed, grip strength, or chair stands, at clinical review. Where these conditions are not met, a detected decline should be classified as unexplained, disease-specific, or situational rather than as hidden frailty. This differential step is built explicitly into Stage 5 of the framework proposed in Section 6.

5.5 Proposed Interpretation of Hidden Frailty

On this basis, we propose an operational definition of hidden frailty: a sustained, multidomain deterioration from an individual's established functional trajectory that may indicate declining physiological reserve before conventional clinical frailty thresholds are reached. This definition is deliberately framed as a hypothesis rather than an established clinical entity, requiring prospective validation against meaningful outcomes before it can be considered a formal diagnostic category. Under this definition, hidden frailty is a property of an individual's measured trajectory, identified using the criteria in Sections 5.3 and 5.4, rather than a stage through which every person passes on the way to prefrailty.

6. Proposed Hidden Frailty Detection Framework

Building on the preceding sections, we propose a six-stage conceptual framework intended to translate the hidden frailty construct into a structured, actionable process, while making explicit the role artificial intelligence is intended to play at each stage. Because each stage is currently supported only by limited or indirect evidence, the framework is presented as a set of testable propositions rather than a validated pathway. Table 4 specifies, for each stage, the candidate measurement variables, provisional decision criteria, clinical response, and the research question that must be answered before the stage can be considered evidence-based, and Figure 1B summarizes the overall pathway.

Table 4. Operationalization of the six-stage framework: candidate variables, provisional criteria, clinical responses, and research priorities.

Stage	Candidate measurement variables	Provisional decision criteria (to be tested)	Clinical response	Priority research question (current evidence)
1. Individualized baseline	Gait speed, cadence, stride variability; daily steps, sedentary bouts, activity fragmentation; sleep efficiency and timing; resting heart rate and HRV; speech or smartphone-interaction features where available	4-8 weeks of data meeting wear-time rules; person-specific mean and variability estimated per domain; refreshed periodically	Conventional frailty assessment (Fried criteria, Frailty Index, or Clinical Frailty Scale) at enrollment to anchor the baseline	What baseline duration yields stable estimates of within-person variability? (Limited; activity volume only)
2. Continuous or repeated monitoring	As Stage 1, plus EHR events (admissions, new prescriptions) and wear-time or adherence data	Pre-specified data-quality rules (e.g., ≥10 h/day on ≥4 days); missingness recorded as a variable	Technical support and adherence prompts; no clinical action	What adherence is sustainable over years, and does it differ by socioeconomic group? (Short-term studies only)
3. Multimodal integration	Domain-level deviation scores	Benchmark: number of domains with reliable deviation; learned models must show incremental value over the benchmark	None (analytic step)	Does multimodal integration outperform the best single domain? (Very limited)
4. Detection of persistent deviation	Reliable change index; CUSUM or EWMA statistics; change-point posterior probability	Deviation beyond reliable-change threshold (e.g., RCI beyond ±1.96), sustained ≥4 weeks, in ≥2 domains, exceeding MCID where available (gait speed ≈ 0.05-0.10 m/s)	Alert sent to a designated clinician or care coordinator	What are the sensitivity, alert rate, and positive predictive value of these criteria? (No prospective evidence)
5. Clinical confirmation	Structured review of pain, acute illness, medication change, mood, deconditioning, social or environmental change; Fried criteria, gait speed, grip strength, chair stands	Alert classified as confirmed decline, alternative explanation, or false alarm	Treat any alternative cause; otherwise proceed to Stage 6	What proportion of alerts reflect reduced reserve rather than other causes? (No evidence)
6. Targeted intervention and follow-up	Response trajectory on the same digital measures; conventional functional outcomes	Return toward baseline within a pre-specified period; escalate if not	Multicomponent exercise, nutrition, medication review, falls and cognitive assessment, rehabilitation	Does signal-triggered intervention improve function compared with usual care? (No evidence; see Section 9)

Note: CUSUM, cumulative sum; EHR, electronic health record; EWMA, exponentially weighted moving average; HRV, heart-rate variability; MCID, minimal clinically important difference; RCI, reliable change index. All thresholds are provisional proposals for prospective testing, not validated criteria.

Stage 1: Establish an individualized baseline. An initial monitoring period characterizes habitual mobility, physical activity, sleep, physiological parameters, and cognitive or behavioral patterns. We provisionally suggest a baseline period of four to eight weeks. Although three to five valid days of accelerometry are sufficient to estimate habitual activity volume in older adults [32], a longer period is needed to characterize each person's day-to-day variability, weekday-weekend patterns, and recovery from minor perturbations, on which individual deviation thresholds depend. Baselines should be refreshed periodically to accommodate seasonal variation and slow, age-expected change.

Stage 2: Continuous or repeated digital monitoring. Ongoing data collection draws on wearable devices, smartphones, ambient sensors, electronic health record data, and other available digital systems, chosen according to feasibility and relevance to a given individual. Minimum data-quality rules, such as the widely used accelerometry criterion of at least 10 hours of valid wear on at least four days [33], should be pre-specified, and periods of non-wear recorded as data in their own right, since declining adherence may itself be informative (Section 6.1).

Stage 3: Multimodal AI integration. Rather than interpreting each data stream independently, an integrative model combines signals across domains, reflecting the multidimensional nature of frailty established earlier in this review. Integration may use simple, interpretable composites, such as the number of domains showing reliable deviation, or learned multimodal models; the simple composite should serve as the benchmark against which any incremental value of more complex models is judged.

Stage 4: Detection of persistent trajectory deviation. The model is tasked specifically with identifying meaningful and sustained deviation from the individual's established baseline, distinguished from transient or artifactual variation.

Provisional criteria for a meaningful deviation are those set out in Section 5.3: change beyond the individual's own reliable-change threshold, persistence beyond a pre-specified period, and involvement of at least two domains.

Stage 5: Clinical confirmation. A detected signal does not itself constitute a diagnosis. It functions as a trigger for appropriate clinical assessment, in which a clinician evaluates the individual and determines whether the signal reflects genuine functional decline. Confirmation should include a structured differential review of alternative explanations, including pain, acute or exacerbated disease, medication change, mood, deconditioning, and environmental or social change (Section 5.4), together with conventional frailty and performance measures, so that each alert is classified as confirmed decline, explained by another cause, or false alarm. Recording these classifications provides the data needed to estimate the positive predictive value of alerts.

Stage 6: Targeted intervention and follow-up. Where clinical confirmation supports genuine deterioration, targeted actions may follow, including structured exercise, nutritional assessment, medication review, falls assessment, cognitive evaluation, and rehabilitation, with continued monitoring to evaluate response. Interventions with evidence of benefit in prefrail and frail older adults, particularly multicomponent exercise and nutritional programs [34], are the natural starting point, although whether they are effective when triggered by a digital signal in people who do not yet meet conventional criteria is itself an untested hypothesis.

The critical message underlying this framework is that artificial intelligence is positioned throughout as an early-warning system rather than an autonomous diagnostic tool. At no stage does the model substitute for clinical judgment; its role is to surface patterns that would otherwise be invisible between clinical encounters and to direct attention toward individuals most likely to benefit from timely assessment. Implementation science offers a useful lens for evaluating whether such a system achieves this in practice: frameworks assessing an intervention's reach, effectiveness, adoption, implementation, and maintenance were developed precisely to move health innovations beyond efficacy demonstrations toward sustained real-world impact, and apply as readily to an AI-enabled monitoring pathway as to any other complex health intervention [35].

Robustness to Missing Data, Non-Adherence, and Transient Illness: The performance of the framework in practice will depend heavily on data completeness, and its sensitivity to realistic rates of missingness has not yet been quantified. Several design choices could reduce this vulnerability. First, analytic methods that tolerate irregular sampling, such as state-space or mixed-effects trajectory models, are preferable to methods that require complete daily series. Second, missingness should be modeled rather than ignored, because non-wear in older adults is unlikely to be random: device abandonment may coincide with illness, hospitalization, cognitive decline, or low mood, so a run of missing data may itself be an early-warning signal rather than mere noise. Third, redundancy across modalities, such as ambient or home sensors that require no wearing, charging, or interaction, can maintain monitoring when a wearable is not used [36]. Fourth, transient illness can be handled by suppressing alerts during, and for a defined recovery window after, documented acute events captured from electronic health records or brief self-report, and then testing whether the trajectory returns to baseline. We recommend that validation studies report alert performance under both observed and simulated missingness of plausible magnitude, and across adherence strata, so that the robustness of each stage is quantified rather than assumed.

7. Limitations of the Current Evidence

7.1 Heterogeneous Frailty Definitions

Studies in this field draw on markedly different operational definitions of frailty, including the Fried phenotype [3], the Frailty Index [4], the Clinical Frailty Scale [6], and numerous modified or setting-specific variants. The consequences of this are not abstract: pooling 240 studies across 62 countries, one meta-analysis found frailty prevalence estimates of 12% using physical-phenotype criteria versus 24% using a deficit-accumulation index in overlapping populations, a twofold discrepancy attributable to instrument choice alone [2], well above even the 6.9% prevalence originally reported in the Cardiovascular Health Study cohort in which the phenotype was first developed [3]. Findings generated using one definition therefore do not necessarily generalize to studies using another, a limitation repeatedly flagged in recent systematic appraisals of frailty-prediction models [7,16].

7.2 Predominance of Cross-Sectional Studies

Across the thirty-one prediction models catalogued in one recent scoping review, 66.7% used cross-sectional designs against roughly 20% using genuinely longitudinal cohorts, though the balance has begun shifting toward longitudinal designs since 2024 [16]. A model that accurately classifies already-established frailty demonstrates only that frailty has recognizable digital correlates, not that those correlates can be detected before frailty becomes clinically apparent, and, as Section 4.2 showed, the classification task is also where reported AUCs are least trustworthy.

These two weaknesses compound one another and bear directly on the central premise of this review. Because prospective, within-person trajectory evidence is almost entirely absent, consisting of a single externally unvalidated model [7], the proposition that clinical frailty is preceded by a detectable digital signal is at present an extrapolation from cross-sectional associations and short-term follow-up, and the cross-sectional models reporting the strongest performance are also those most affected by feature-outcome overlap (Section 7.4). A cross-sectional association between a digital feature and current

frailty status does not establish that within-person change in that feature precedes, or predicts, transition to frailty; establishing this temporal ordering requires repeated measurement in individuals who are robust or prefrail at enrollment, with follow-up to incident frailty or functional decline. Until such evidence exists, hidden frailty should be regarded as a hypothesis that the existing literature motivates but does not directly support, and the framework in Section 6 as a design for testing that hypothesis rather than a description of a validated phenomenon. To keep the concept falsifiable, we state in advance what would count against it: if, in prospective cohorts, within-person digital deviation does not anticipate incident frailty or functional transition earlier or more accurately than repeated conventional assessment, or does so only through predictors that form part of the outcome definition, the concept would add little clinical value.

7.3 Small, Selective Datasets and Limited External Validation

Development cohorts in this literature range enormously, from 1,230 to 86,133 participants in one systematic review [7], and many remain constrained by single-center recruitment and highly selected, research-engaged volunteers who differ systematically from the broader population of older adults [16]. External validation is the exception rather than the rule: only 22.6% of the thirty-one models in the scoping review above were tested outside their development data, and validation that did occur was "mostly confined to the same country or database, without independent cross-population testing" [16]. A parallel systematic review applying the PROBAST risk-of-bias tool found that only 2 of 14 studies achieved both low risk of bias and low applicability concerns [7], meaning the great majority of published models cannot currently be judged reliable even on their own authors' reported terms. Because that review did not report the performance of its two low-risk studies separately [7], the effect of restricting the evidence to low-risk studies cannot be quantified from published data. Given that the highest classification AUCs were attributed to circularity, such a restriction would most likely shift the plausible performance range downward toward that of prospective prediction models, and would leave the trajectory-detection question with almost no supporting evidence.

7.4 Risk of Circular Prediction

Where the features used to train a model overlap substantially with the criteria used to define the frailty outcome itself, for instance, using gait speed to predict a frailty phenotype that itself incorporates gait speed, apparently excellent model performance may simply reflect this construct overlap rather than genuine predictive insight. This is not a hypothetical concern: it is the explanation one systematic review offered for its own field's highest reported classification accuracies, AUCs of 0.97-0.98 [7], a risk under-acknowledged in much of the remaining literature. A precise estimate of the inflation attributable to feature-outcome overlap cannot be derived from the published literature, because no study has compared the same model with and without overlapping predictors. An indirect upper bound is given by the gap between the circularity-prone classification models (AUC 0.97-0.98) and prospective prediction models (internal AUC 0.70-0.81) [7], a difference of roughly 0.16-0.28; this gap, however, conflates circularity with the inherently greater difficulty of forecasting future status and should not be read as a direct estimate. We recommend that future studies report performance with and without predictors that form part of the outcome definition as a routine sensitivity analysis.

7.5 Overreliance on Discrimination Metrics

Model evaluation in this field has relied heavily on discrimination statistics such as the AUC, an insufficient basis for judging clinical usefulness alone [37]. The gap is measurable: one scoping review found that some calibration metric was reported for 74.2% of models overall but for only 50% of machine-learning models specifically [16], meaning the more complex methods that most need calibration scrutiny receive it least often. Predictive values under realistic prevalence, decision-threshold utility, and the anticipated burden of false positives and negatives require far greater attention than they currently receive, attention that current reporting guidance for clinical prediction models explicitly calls for [38].

7.6 Regulatory and Implementation Gaps

Few models in this field have progressed anywhere near clinical deployment. One systematic review, using standard technology-readiness-level criteria, found the literature concentrated at TRL 4-6, with no studies reaching TRL 7 or higher and none documenting the implementation or maintenance domains of established implementation-science frameworks [7,35]. Frameworks such as the US Food and Drug Administration's approach to AI- and machine-learning-based software as a medical device envisage iterative development with planned post-market performance monitoring, an expectation this literature has barely begun to engage with [39].

7.7 Limited Evidence of Clinical Benefit

The most consequential unresolved question is not whether digital biomarkers and artificial intelligence can detect subtle functional signals, evidence increasingly suggests that they can, but whether acting on those signals actually prevents or delays adverse functional outcomes. This is not merely an unaddressed question but, on the evidence assembled above, close to entirely unaddressed: the same scoping review of thirty-one models found no studies conducting health-economic evaluation or prospective validation of clinical impact whatsoever [16]. No amount of predictive accuracy resolves this question; only prospective intervention studies can, and such studies remain absent from the current evidence base. The absence of economic evidence is particularly consequential for health systems with limited resources, in which continuous

monitoring would compete with established services for funding. Economic evaluations should capture not only device and analytic costs but also the clinician time required to review alerts, the downstream costs of false positives, and potential savings from avoided falls, admissions, and loss of independence; until such data exist, the short-term affordability of the framework, particularly in publicly funded and lower-resource systems, cannot be assumed.

7.8 Selection, Publication, and Reporting Bias

The primary literature is likely to be affected by publication and selective-reporting bias. Small, single-cohort studies reporting high AUCs, such as several of those in Table 1, are more likely to be published than null or modest results, a pattern well documented in the prognostic literature more broadly [40], and selective reporting of the best-performing of several tested models is common in machine-learning studies. Neither of the principal syntheses on which this review draws assessed publication bias formally: one did not undertake a meta-analysis [7], and the other did not undertake any formal risk-of-bias assessment [16]. Formal tests for small-study effects were not appropriate in the present review because estimates were not pooled, and the heterogeneity of outcomes, populations, and metrics would make such tests uninformative. The high single-cohort AUCs discussed in Section 3 should therefore be read as optimistic estimates.

7.9 Limitations of this Review

This review has limitations of its own. As a narrative rather than systematic review, it did not use a registered protocol, duplicate screening, or exhaustive searching, and its selection of studies, despite the structured searches and explicit prioritization criteria described in Section 1.1, may be subject to selection bias. Quality appraisal of primary studies was structured but narrative (Table 2) rather than a formal duplicate PROBAST assessment, and no quantitative pooling was undertaken. Several of the cited studies were published recently and have not yet been independently replicated. These features suit the conceptual aim of the review, which is to propose and bound a testable framework, but they mean that the quantitative estimates reported here should be regarded as illustrative of the current evidence rather than as definitive summary measures. A systematic review with formal risk-of-bias assessment would be a valuable next step once more trajectory-based studies become available.

8. Equity, Ethics, and Clinical Implementation

8.1 Digital Inequity

Access to the wearable devices, smartphones, and connected infrastructure underlying digital frailty monitoring is not distributed evenly across the older adult population. Income, geographic location, technological literacy, and disability all shape whether an individual can meaningfully participate in continuous digital monitoring, and these access gaps have been shown to persist, and in some analyses widen, along lines of race, income, and rurality even as overall internet and smartphone adoption among older adults has grown [41]. Because the population most likely to benefit from earlier detection is also aging fastest in lower-resource regions where digital infrastructure and geriatric services are least developed [1], a detection system built and validated predominantly among digitally engaged, higher-resource populations risks widening rather than narrowing existing disparities in geriatric care. Evidence that the digital signals reviewed here can be captured with comparable accuracy and completeness across socioeconomic and geographic strata is currently sparse. Most validation cohorts are drawn from high-income settings and research-engaged volunteers, and we identified no study that formally compared the accuracy or completeness of frailty-relevant digital signals across income, education, or rural-urban strata. Measurement accuracy may itself vary systematically: some step counters are least accurate at the slow walking speeds typical of frail older adults [42], optical heart-rate sensors show accuracy that varies with activity type [43], and pulse oximetry has been shown to overestimate oxygen saturation more often in Black than in White patients [44]. Validation studies should therefore report measurement accuracy, data completeness, and model performance stratified by socioeconomic position, geography, ethnicity, and functional level.

Stratified reporting is necessary for detecting inequity, but it is not sufficient to prevent it: it reveals differences in performance without changing who is able to take part. Structural barriers, including the cost of devices and data plans, unreliable connectivity, the need to charge and wear a device daily, limited digital literacy, and cognitive impairment, are likely to reduce adherence precisely among older adults in low-resource settings and those with cognitive decline, who may have the most to gain from earlier detection. The framework therefore needs to be designed around these barriers rather than evaluated against them after the fact. Practical measures include providing devices and connectivity at no cost to participants, as part of the intervention rather than as a precondition for it; prioritizing low-burden modalities that require little or no daily action, such as long-battery-life devices, passive ambient sensors, and routinely collected electronic-health-record or primary-care data; supporting use through family members, caregivers, or community health workers where the person wishes; storing data locally and transmitting it opportunistically where connectivity is intermittent; and adapting the analysis so that sparse or interrupted data reduce the confidence attached to an alert rather than excluding the person from monitoring altogether. Development and validation cohorts should deliberately include participants from lower-resource settings and people with cognitive impairment, and adherence and reach should be reported as equity outcomes in their own right. A detection pathway should also not be deployed where there is no capacity to respond to its alerts, because unequal access to follow-up services would convert earlier detection into unequal benefit.

Even with these measures, the framework could widen disparities if it is adopted first, or only, in well-resourced systems; its equity impact should therefore be assessed prospectively, including in the evaluation trial described in Section 9.

8.2 Algorithmic Bias

Even where access is equivalent, algorithmic performance may not be. Differences in model accuracy across age, sex, ethnicity, socioeconomic context, and healthcare setting have been documented across numerous clinical AI applications, most notably in a widely cited demonstration that a commercial population-health algorithm systematically underestimated the care needs of Black patients relative to White patients of equivalent underlying illness because it used healthcare cost as a proxy for health need [45], and there is no reason to assume frailty-detection models are immune.

8.3 Privacy, Surveillance, and Acceptability

Passive, continuous monitoring is, by its nature, intrusive in ways that periodic clinical assessment is not. Data streams capable of revealing hidden frailty are equally capable of revealing an individual's movements, daily routines, sleep patterns, and social behavior, a tension that qualitative work with older adults themselves has found is felt acutely: many express willingness to accept in-home monitoring in principle while remaining concerned about who ultimately sees the data and how it might be used [46]. A related but distinct question concerns the acceptability of assistive and companionship technologies more broadly, including socially assistive robots, whose benefits for engagement and mood in dementia care have been demonstrated in controlled trials but whose appropriate role alongside passive monitoring remains contested [47]. Both raise substantive questions of informed consent and data governance that require deliberate ethical and regulatory attention rather than being treated as secondary to technical development.

These concerns are amplified for older adults with mild cognitive impairment or early dementia, the group for whom unobtrusive ambient monitoring is most often proposed as an alternative to wearables. In-home sensors and cameras may be experienced as surveillance and as a loss of control over one's own home, may provoke anxiety or a sense of being distrusted, and may carry stigma when visible to visitors; older adults' acceptance of sensor monitoring has been found to depend on perceived need, on control over who sees the data, and on the avoidance of cameras in private spaces [46,48]. Capacity to consent may fluctuate or decline over the very period that monitoring is intended to cover, so consent should be treated as an ongoing process rather than a one-off event, supported decision-making and, where appropriate, advance or proxy agreement should be used, and monitoring should be easy to pause or withdraw. Technical design can also reduce intrusiveness: non-visual sensors such as motion, door, and bed sensors [36] or radar in place of cameras; on-device processing that transmits only derived features rather than raw images or audio; data minimization; and transparent displays that allow individuals and families to see what is collected. Psychological acceptability should be measured as an outcome in its own right, and adverse effects such as distress or behavioral change under monitoring should be recorded in trials. Without such safeguards, the population that might benefit most from ambient monitoring could also be the one least able to decline it.

8.4 Clinical and Regulatory Integration

Finally, a functioning system requires clear answers to practical implementation questions frequently left unaddressed in the technical literature: who receives an alert; what threshold of trajectory deviation triggers action; what specific intervention follows a confirmed signal; who is accountable for false alarms; and, increasingly, what regulatory pathway and post-market surveillance obligations apply once such a system leaves the research setting [39]. Co-designing these systems with older adults themselves, rather than only with clinicians and engineers, has been proposed as one route to systems that are both acceptable and genuinely useful in daily life [49].

9. Future Research Priorities

Advancing this field requires specific, actionable research priorities rather than generic calls for further study.

Prospective longitudinal cohorts. Studies should recruit robust and prefrail older adults specifically, rather than those already frail, and follow them prospectively with repeated multidomain measurement rather than a single baseline and endpoint assessment.

Individualized trajectory models. Analytic approaches should prioritize within-person change over time rather than between-person comparison at a single point, and should report performance in terms of trajectory-detection accuracy rather than cross-sectional classification accuracy alone.

Multimodal digital biomarkers. Future work should combine physiological and behavioral domains rather than continuing to evaluate single data streams in isolation, reporting the incremental value each added domain contributes.

External validation. Models require systematic testing across countries, ethnic groups, healthcare systems, and socioeconomic contexts, ideally through pre-registered validation studies conducted independently of the original development team.

Standardization. The field would benefit substantially from consensus definitions of early functional decline, clinically meaningful change, and standardized digital frailty endpoints, developed through structured collaboration among geriatricians, data scientists, and patient representatives.

Clinical utility and implementation studies. Research emphasis should shift from model performance metrics toward intervention effectiveness and real-world implementation, evaluated through randomized or stepped-wedge trial designs and assessed against the reach, adoption, and maintenance dimensions relevant to any complex health intervention [35].

Reporting, regulatory, and co-design standards. The field should adopt rigorous reporting standards such as TRIPOD+AI [38], engage regulatory frameworks for AI-based software as a medical device early in development rather than retrospectively [39], and involve older adults directly in co-designing systems intended for their own use [49].

Economic evaluation. Model-based and trial-based cost-effectiveness analyses should be undertaken early, including in lower-resource health systems, and should compare low-cost configurations, such as smartphone-only or electronic-health-record-based monitoring, with multisensor configurations, since the least expensive configuration that achieves adequate detection is the one most likely to be scalable within public health systems.

Definitive evaluation trial. Moving hidden frailty from a conceptual framework to an evidence-based clinical pathway will require a pragmatic randomized trial, preferably cluster-randomized at the level of primary-care practices or community programs, or of stepped-wedge design where phased implementation is planned. Participants would be community-dwelling adults aged 70 years or older who are robust or prefrail at enrollment. The intervention arm would receive the full pathway (continuous monitoring, alert generation, structured clinical confirmation, and multicomponent intervention); the control arm would receive usual care, ideally with identical monitoring devices but suppressed alerts, which separates the effect of acting on signals from the effect of wearing devices and allows the accuracy of the alerts that would have been generated to be assessed. We suggest a patient-centered primary outcome, such as time to incident disability in basic activities of daily living or days alive and living independently at home, assessed over at least 24 months. Secondary outcomes should include incident frailty on a pre-specified instrument, falls, hospitalization, quality of life, alert positive predictive value, adherence, psychological acceptability, cost per quality-adjusted life-year, and outcomes stratified by socioeconomic group. A feasibility study and prospective observational validation of the detection criteria in Table 4 would be needed first to inform the sample-size calculation.

10. Clinical and Public Health Implications

The implications outlined in this section are conditional. At present, no study has shown that acting on a digitally detected signal of early decline improves outcomes, no health-economic evaluation exists, and the trajectory-based detection on which the framework depends has been attempted by a single, externally unvalidated model [7,16]. What follows should therefore be read as a description of what validation would need to demonstrate, not as a forecast of benefit.

Should the hidden frailty framework be validated through the prospective research outlined above, digitally enabled detection systems could plausibly support primary care, specialist geriatric services, community aging programs, rehabilitation, and home-based monitoring, extending the reach already demonstrated for telehealth more broadly in improving access and outcomes for older adults living independently [50]. In each of these settings, the same basic pathway would apply: an artificial intelligence system identifies emerging deterioration, a clinician confirms the underlying cause, a targeted intervention is implemented, and longitudinal monitoring continues to evaluate response. Given how often frailty coexists with multimorbidity and polypharmacy, both of which independently shape functional trajectory and complicate the interpretation of any single deviation [18,51], meaningful implementation will require integration with, rather than addition to, existing chronic-disease management pathways.

The public health significance of this pathway lies not in its capacity to generate a risk score, but in its potential to preserve function that would otherwise be lost. If a system were shown to identify deterioration reliably, weeks or months earlier than conventional assessment, it could open a window during which intervention may be more effective because functional reserve has not yet been fully depleted; this premise is plausible but has not yet been demonstrated. The value proposition of hidden frailty detection is therefore best understood in terms of preserved independence, sustained mobility, maintained resilience, and protected quality of life, rather than in terms of algorithmic accuracy or risk stratification for its own sake. Realizing this potential at a population level would require integration with existing care pathways, sufficient workforce capacity to respond to generated alerts, and sustained investment in infrastructure that reaches the older adult population equitably, including in the lower-resource settings where population aging is proceeding fastest and where specialist workforce and service capacity for age-related conditions often remain limited [1,52].

In resource-constrained public health systems, a staged, lower-cost approach is more realistic than universal multisensor monitoring: for example, using routinely collected electronic-health-record data or smartphone-based measures to identify individuals for targeted, time-limited monitoring, and reserving clinician review for persistent multidomain alerts. Whether any such configuration is cost-effective remains to be established.

11. Conclusion

This review has argued that frailty is best understood as a dynamic process rather than a static classification, and that conventional assessment tools, while clinically valuable and well validated, are structurally limited to episodic snapshots that can only detect deterioration once it has already accumulated to a recognizable threshold. Advances in wearable technology, smartphones, ambient sensing, electronic health records, computer vision, and machine learning create a genuine, if still largely unproven, opportunity to characterize subtle, multidomain longitudinal changes occurring before conventional frailty becomes clinically apparent, a state we have termed hidden frailty.

Realizing this opportunity depends on three conceptual transitions running throughout this review: from static classification toward trajectory detection, from population-derived thresholds toward individualized functional baselines, and from algorithmic accuracy alone toward clinically actionable early warning embedded within responsible, adequately regulated care pathways. None of these transitions is complete in the current literature, and each faces substantial methodological, ethical, regulatory, and implementation barriers detailed in the preceding sections.

Hidden frailty, as presented here, remains a testable conceptual framework rather than an established clinical entity. Its ultimate importance will not be determined by predictive accuracy or technical sophistication, but by whether earlier identification, achieved responsibly and equitably across the populations who stand to benefit most, translates into effective interventions and demonstrably better functional outcomes for older adults. The most immediate priorities are prospective validation of the proposed detection criteria, transparent reporting of calibration and clinical utility, and a pragmatic trial testing whether signal-triggered intervention improves functional outcomes.

Acknowledgements

None.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethics Statement

This article is a review of previously published literature and did not involve human participants, human data, or animals; ethical approval and informed consent were therefore not required.

Informed Consent

Not applicable. This article is a review of previously published literature and did not involve human participants.

Data Availability Statement

No new data were generated or analyzed in this study. All data discussed are available in the published sources cited.

Author Contributions

All authors (A.D.A., U.U.O., and I.A.O.) contributed equally to the conceptualization, literature search, analysis and interpretation of the evidence, writing of the original draft, and review and editing of the manuscript. All authors read and approved the final manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

Generative AI Statement

During the preparation of this manuscript, the authors used Claude (Anthropic) solely for language editing and polishing. The authors reviewed all text and take full responsibility for the content of this publication.

References

[1] United Nations, Department of Economic and Social Affairs, Population Division. World population prospects 2024: Summary of results (UN DESA/POP/2024/TR/NO. 9). New York: United Nations. 2024.

[2] O'Caoimh R, Sezgin D, O'Donovan MR, Molloy DW, Clegg A, Rockwood K, et al. Prevalence of frailty in 62 countries across the world: A systematic review and meta-analysis of population-level studies. Age Ageing, 2021, 50(1), 96-104. DOI: 10.1093/ageing/afaa219
[3] Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: Evidence for a phenotype. The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 2001, 56(3), M146-M156. DOI: 10.1093/gerona/56.3.M146
[4] Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 2007, 62(7), 722-727. DOI: 10.1093/gerona/62.7.722
[5] Searle SD, Mitnitski A, Gahbauer EA, Gill TM, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatrics, 2008, 8, 24. DOI: 10.1186/1471-2318-8-24
[6] Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I, et al. A global clinical measure of fitness and frailty in elderly people. Canadian Medical Association Journal, 2005, 173(5), 489-495. DOI: 10.1503/cmaj.050051
[7] Kim S, Shin MJ, Choi BK, Obradovic Z, Rubin DJ, Park JH. Machine learning-based frailty prediction and classification in community-dwelling older adults: A systematic review of validation, explainability, and implementation readiness. Healthcare (Basel), 2026, 14(11), 1543. DOI: 10.3390/healthcare14111543
[8] Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Research Integrity and Peer Review, 2019, 4, 5. DOI: 10.1186/s41073-019-0064-8
[9] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 2019, 170(1), 51-58. DOI: 10.7326/M18-1376
[10] Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. The Journal of the American Medical Association, 2011, 305(1), 50-58. DOI: 10.1001/jama.2010.1923
[11] Schwenk M, Mohler J, Wendel C, D'Huyvetter K, Fain M, Taylor-Piliae R, et al. Wearable sensor-based in-home assessment of gait, balance, and physical activity for discrimination of frailty status: Baseline results of the Arizona frailty cohort study. Gerontology, 2015, 61(3), 258-267. DOI: 10.1159/000369095
[12] Mou C, Yan X, Miao X, Zhu L. Accuracy of wearable devices in predicting falls in older adults: A systematic review and meta-analysis. Frontiers in Public Health, 2026, 14, 1778750. DOI: 10.3389/fpubh.2026.1778750
[13] Culos A, Manas A, Shidara K, Fallazadeh R, Garcia-Garcia FJ, Losa-Reyna J, et al. A machine learning model for frailty based on wearable device measurements. Communications Medicine, 2026, 6(1), 173. DOI: 10.1038/s43856-026-01419-7
[14] Varadhan R, Chaves PH, Lipsitz LA, Stein PK, Tian J, Windham BG, et al. Frailty and impaired cardiac autonomic control: New insights from principal components aggregation of traditional heart rate variability indices. The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 2009, 64(6), 682-687. DOI: 10.1093/gerona/glp013
[15] Da Cunha E, Zory R, Chorin F, Manera V, Gros A. Spontaneous speech enables scalable digital phenotyping of physical functional deficits in aging. NPJ Aging, 2026, 12(1), 52. DOI: 10.1038/s41514-026-00343-3
[16] Xiao J, Zhang W, Xu D, Mao H, Yang H. Predictive risk models for frailty onset in older adults: A scoping review of methodological trends, model performance, and clinical translation gap. Clinical Interventions in Aging, 2026, 21, 620467. DOI: 10.2147/CIA.S620467
[17] Xue QL. The frailty syndrome: Definition and natural history. Clinics in Geriatric Medicine, 2011, 27(1), 1-15. DOI: 10.1016/j.cger.2010.08.009
[18] Vetrano DL, Palmer K, Marengoni A, Marzetti E, Lattanzio F, Roller-Wirnsberger R, et al. Frailty and multimorbidity: A systematic review and meta-analysis. The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 2019, 74(5), 659-666. DOI: 10.1093/gerona/gly110
[19] Gill TM, Gahbauer EA, Allore HG, Han L. Transitions between frailty states among community-living older persons. Archives of Internal Medicine, 2006, 166(4), 418-423. DOI: 10.1001/archinte.166.4.418
[20] Cesari M, Araujo de Carvalho I, Amuthavalli Thiyagarajan J, Cooper C, Martin FC, Reginster JY, et al. Evidence for the domains supporting the construct of intrinsic capacity. The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 2018, 73(12), 1653-1660. DOI: 10.1093/gerona/gly011
[21] Benkaci A, Sliman L, Dellys HN. Vision-based human fall detection systems: A review. Procedia Computer Science, 2024, 241, 203-211. DOI: 10.1016/j.procs.2024.08.028
[22] Ensrud KE, Blackwell TL, Redline S, Ancoli-Israel S, Paudel ML, Cawthon PM, et al. Sleep disturbances and frailty status in older community-dwelling men. Journal of the American Geriatrics Society, 2009, 57(11), 2085-2093. DOI: 10.1111/j.1532-5415.2009.02490.x
[23] Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology, 2016, 41(7), 1691-1696. DOI: 10.1038/npp.2016.7
[24] Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digital Medicine, 2020, 3, 119. DOI: 10.1038/s41746-020-00323-1
[25] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Medical Informatics and Decision Making, 2020, 20(1), 310. DOI: 10.1186/s12911-020-01332-6
[26] Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of Clinical Epidemiology, 2015, 68(1), 25-34. DOI: 10.1016/j.jclinepi.2014.09.007
[27] Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Medical Decision Making, 2006, 26(6), 565-574. DOI: 10.1177/0272989X06295361
[28] Jacobson NS, Truax P. Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. Journal of Consulting and Clinical Psychology, 1991, 59(1), 12-19. DOI: 10.1037/0022-006X.59.1.12
[29] Perera S, Mody SH, Woodman RC, Studenski SA. Meaningful change and responsiveness in common physical performance measures in older adults. Journal of the American Geriatrics Society, 2006, 54(5), 743-749. DOI: 10.1111/j.1532-5415.2006.00701.x
[30] Folorunsho S, Lawal OP, Somuah MA. Functional impairment without a depression diagnosis as an indicator of psychological distress in older adults. Cambridge Prisms: Global Mental Health, 2025, 12, e138. DOI: 10.1017/gmh.2025.10098

[31] Gijzel SMW, Whitson HE, van de Leemput IA, Scheffer M, van Asselt D, Rector JL, et al. Resilience in clinical care: Getting a grip on the recovery potential of older adults. Journal of the American Geriatrics Society, 2019, 67(12), 2650-2657. DOI: 10.1111/jgs.16149
[32] Hart TL, Swartz AM, Cashin SE, Strath SJ. How many days of monitoring predict physical activity and sedentary behaviour in older adults? The International Journal of Behavioral Nutrition and Physical Activity, 2011, 8, 62. DOI: 10.1186/1479-5868-8-62
[33] Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Medicine & Science in Sports & Exercise, 2008, 40(1), 181-188. DOI: 10.1249/mss.0b013e31815a51b3
[34] Apóstolo J, Cooke R, Bobrowicz-Campos E, Santana S, Marcucci M, Cano A, et al. Effectiveness of interventions to prevent pre-frailty and frailty progression in older adults: A systematic review. JBI Database of Systematic Reviews and Implementation Reports, 2018, 16(1), 140-232. DOI: 10.11124/JBISRIR-2017-003382
[35] Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. American Journal of Public Health, 1999, 89(9), 1322-1327. DOI: 10.2105/AJPH.89.9.1322
[36] Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems for assessing aging changes: Home-based, unobtrusive, and continuous assessment of aging. The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 2011, 66B(Suppl 1), i180-i190. DOI: 10.1093/geronb/gbq095
[37] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: The Achilles heel of predictive analytics. BMC Medicine, 2019, 17(1), 230. DOI: 10.1186/s12916-019-1466-7
[38] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. British Medical Journal, 2024, 385, e078378. DOI: 10.1136/bmj-2023-078378
[39] US Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan. Silver Spring, MD: FDA, 2021.
[40] Kyzas PA, Denaxa-Kyza D, Ioannidis JPA. Almost all articles on cancer prognostic markers report statistically significant results. European Journal of Cancer, 2007, 43(17), 2559-2579. DOI: 10.1016/j.ejca.2007.08.030
[41] Yang R, Gao S, Jiang Y. Digital divide as a determinant of health in the U.S. older adults: Prevalence, trends, and risk factors. BMC Geriatrics, 2024, 24(1), 1027. DOI: 10.1186/s12877-024-05612-y
[42] Grant PM, Dall PM, Mitchell SL, Granat MH. Activity-monitor accuracy in measuring step number and cadence in community-dwelling older adults. Journal of Aging and Physical Activity, 2008, 16(2), 201-214. DOI: 10.1123/japa.16.2.201
[43] Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digital Medicine, 2020, 3, 18. DOI: 10.1038/s41746-020-0226-6
[44] Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. The New England Journal of Medicine, 2020, 383(25), 2477-2478. DOI: 10.1056/NEJMc2029240
[45] Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 2019, 366(6464), 447-453. DOI: 10.1126/science.aax2342
[46] Mheidly N, Dempsey PR, Quinn K. Older adults' privacy perceptions of passive in-home monitoring technologies: A comprehensive scoping review. Ageing & Society, 2026, 46, e58. DOI: 10.1017/S0144686X26100701
[47] Lu LC, Lan SH, Hsieh YP, Lin LY, Lan SJ, Chen JC. Effectiveness of companion robot care for dementia: A systematic review and meta-analysis. Innovation in Aging, 2021, 5(2), igab013. DOI: 10.1093/geroni/igab013
[48] Pol M, van Nes F, van Hartingsveldt M, Buurman B, de Rooij S, Kröse B. Older people's perspectives regarding the use of sensor monitoring in their home. Gerontologist, 2016, 56(3), 485-493. DOI: 10.1093/geront/gnu104
[49] Sumner J, Chong LS, Bundele A, Lim YW. Co-designing technology for aging in place: A systematic review. Gerontologist, 2021, 61(7), e395-e409. DOI: 10.1093/geront/gnaa064
[50] Fernandes G, Figueiredo T, Costa E, Coelho L, Loyens D. Impact of telehealth on health outcomes and quality of life in the older adults population: A systematic review. Frontiers in Digital Health, 2025, 7, 1708960. DOI: 10.3389/fdgth.2025.1708960
[51] Palmer K, Villani ER, Vetrano DL, Cherubini A, Cruz-Jentoft AJ, Curtin D, et al. Association of polypharmacy and hyperpolypharmacy with frailty states: A systematic review and meta-analysis. European Geriatric Medicine, 2019, 10(1), 9-36. DOI: 10.1007/s41999-018-0124-5
[52] Elechi KW, Soyemi T, Shekoni M, Ede NS, Elechi US, Iloanusi CA, et al. Building healthcare capacity for neurodegenerative disease management in Nigeria: Challenges and opportunities. Journal of Public Health Research, 2025, 14(2). DOI: 10.1177/22799036251350957