BMC Geriatrics

SYSTEMATIC REVIEW Open Access

Risk prediction models for falls in hospitalized older patients: a systematic review and meta-analysis

Anli Mao[1], Jie Su[1], Mingzhu Ren[1], Shuying Chen[1] and Huafang Zhang[1*]

Abstract

Background Existing fall risk assessment tools in clinical settings often lack accuracy. Although an increasing number of fall risk prediction models have been developed for hospitalized older patients in recent years, it remains unclear how useful these models are for clinical practice and future research.

Objectives To systematically review published studies of fall risk prediction models for hospitalized older adults.

Methods A search was performed of the Web of Science, PubMed, Cochrane Library, CINAHL, MEDLINE, and Embase databases: to retrieve studies of predictive models related to falls in hospitalized older adults from their inception until January 11, 2024. Extraction of data from included studies, including study design, data sources, sample size, predictors, model development and performance, etc. Risk of bias and applicability were assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) checklist.

Results A total of 8086 studies were retrieved, and after screening, 13 prediction models from 13 studies were included. Four models were externally validated. Eight models reported discrimination metrics and two models reported calibration metrics. The most common predictors of falls were mobility, fall history, medications, and psychiatric disorders. All studies indicated a high risk of bias, primarily due to inadequate study design and methodological flaws. The AUC values of 8 models ranged from 0.630 to 0.851.

Conclusions In the present study, all included studies had a high risk of bias, primarily due to the lack of prospective study design, inappropriate data analysis, and the absence of robust external validation. Future studies should prioritize the use of rigorous methodologies for the external validation of fall risk prediction models in hospitalized older adults.

Trial registration The study was registered in the International Database of Prospectively Registered Systematic Reviews (PROSPERO) CRD42024503718.

Keywords Risk prediction model, Falls, Hospitalized older adult, Aged, Systematic review, Meta-analysis

*Correspondence:
Huafang Zhang
wuyizhangyichi@zju.edu.cn
[1]Department of Nursing, the Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu 322000, China

Background

Falls are a significant global concern, resulting in 684,000 deaths annually, according to the World Health Organization [1]. Falls represent a leading cause of disability among older adults, posing a significant problem even for those in good health. The growing older adults and increasing life expectancy make fall prediction increasingly important. Hospital-acquired falls (HAFs) are a particular concern for healthcare systems [2], with roughly 28% of hospitalized patients reporting a fall within the past year and 15% experiencing one during their stay [3]. It is understood that approximately 1–3% of hospitalized patients who experience falls may suffer from fractures as a result [4]. In addition, falls may also lead to subdural hematomas and hemorrhages, which not only have a significant impact on the health and quality of life of older adults but also place a heavy burden on families and the healthcare system.

Despite a focus on fall reduction in many studies, current fall risk assessment tools and evidence-based practices have limitations in effectiveness [5, 6]. This includes the potential for a time-consuming assessment process and the influence of subjective judgments by healthcare professionals. Moreover, these assessment tools typically rely on static risk factors and fail to account for the dynamic changes in patients' conditions during their hospital stay. Therefore, a highly accurate and easy-to-use tool is crucial for identifying fall risks in hospitalized older adults. Additionally, translating research findings into clinical practice is essential to enhance safety for hospitalized older adult patients [7].

In recent years, artificial intelligence (AI) has been playing an increasingly important role in medical diagnosis by analyzing medical records, exams, and test results to identify disease patterns and improve diagnostic accuracy [8]. Prediction models are a significant branch of artificial intelligence and serve as a vital quantitative tool for assessing clinical risks and benefits. However, despite the increasing number of prediction models for fall risk in hospitalized older adults, they commonly face several key challenges, including insufficient data quantity, limitations in clinical validation, and a lack of adaptability to different patient populations. These issues restrict the widespread application of these models in clinical practice. Our study aimed to conduct a systematic assessment of these models, integrate the evidence pertaining to risk factors for falls among hospitalized older adults, and provide valuable references for future research and clinical practice.

Methods

Design

Following the established guidelines for evaluating predictive models [9] and the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [10], we conducted a systematic review. The protocol for this review was prospectively registered on the PROSPERO International Prospective Register of Systematic Reviews website (CRD42024503718).

Search strategy

We conducted a comprehensive search of multiple databases and search platforms, including Web of Science, PubMed, Cochrane Library, CINAHL, MEDLINE, and Embase, from their inception until January 11, 2024, that investigated fall risk prediction models in hospitalized older adults aged 65 and older. We also conducted a manual review of the references from the retrieved studies. Our search utilized a combination of medical subject headings (MeSH) and text words, incorporating the following four concepts: (1) inpatients, inpatient, hospital*; (2) aged, elderly, senium, older adults, senior citizen; (3) accidental falls, fall, falling; (4) prediction model, risk score, risk assessment, risk prediction. A complete list of search terms is available in Appendix A. A detailed description of the population, interventions, comparisons, outcomes, timing, and settings (PICOTS) for this systematic review is provided below:

P (Population): ≥65 years old hospitalized older patients.

I (Intervention): Risk prediction models for falls.

C (Comparator): Not applicable.

O (Outcome): Presence of fall.

T (Timing): During the hospitalization.

S (Setting): Hospitalized patients only.

Inclusion and exclusion criteria

To be included in this review, studies had to meet the following criteria: (1) participants were hospitalized patients aged 65 years or older, (2) the study design was observational, (3) the study developed and/or validated a multivariable predictive model with at least two predictors of falls, and (4) the primary outcome of interest was falls during hospitalization. Studies were excluded if they did not meet any of the following criteria: (1) Falls were assessed using an assessment scale, (2) they used a cross-sectional survey design, (3) the outcome measure focused on adverse events due to falls rather than falls themselves, (4) the language of the study was not English, or (5) the full text of the article was not available.

Study selection

Duplicate records were removed using Zotero software. Two independent reviewer pairs (MAL and RMZ) screened titles and abstracts against the inclusion/exclusion criteria for fall prediction model studies. Disagreements were resolved through discussion, with a third

reviewer (CSY) consulted when needed. After reaching a consensus, two reviewers (MAL and SJ) independently screened full texts. Additionally, reference lists of included studies were reviewed for potentially relevant articles.

Data extraction

Two reviewers(MAL and SJ) independently extracted data based on the critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) [10]. Extracted information included basic details like authors, publication year, study design, participants, data sources, and sample size. Specific to predictive modeling, we extracted details on variable selection methods, model development techniques, validation types, performance measures, handling of missing data and continuous variables, predictors used in the final model, and the model presentation format. For studies with multiple models, we focused on the one with the best predictive performance. Any disagreements in data extraction were resolved through discussion (MAL, SJ, and ZHF).

Quality assessment

To assess the risk of bias (ROB) and applicability of prediction models in the included studies, we utilized the Prediction Model Risk of Bias Assessment Tool (PROBAST) [11]. This tool features 20 key questions across four domains: study population, predictors, outcomes, and statistical analysis. The first three domains assess applicability, similar to the Risk of Bias tool but excluding specific risk of bias questions. Each question has answer options like "yes", "probably yes", "no", "probably no", or "no information". A domain is considered high risk if it has at least one "no" or "probably no" answer. If one or more domains are unclear and the others are low risk, the overall bias is unclear. Overall low risk of bias requires all domains to be judged low risk. Two authors (MAL and SJ) independently assessed quality using PROBAST. In case of disagreements regarding quality assessment, a discussion involving three authors (MAL, SJ, and ZHF) was held to reach a consensus.

Data analysis

When more than two studies reported the same outcome measure, a meta-analysis was performed. We used the 'metamisc' package in R software (version 4.2.3) to estimate unreported AUC confidence intervals and calculate predictive intervals. The random-effects model with the Hartung-Knapp-Sidik-Jonkman (HKSJ) method was used to calculate the 95% confidence interval for the average performance (The HKSJ method can provide more accurate type I error rates and confidence intervals when heterogeneity exists) [12, 13]. Heterogeneity was estimated using the predictive intervals calculated by the HKSJ method, with a wider predictive interval compared to the confidence interval indicating the presence of heterogeneity among the original studies [14]. To investigate the sources of variation, subgroupings of different modelling approaches were performed. Sensitivity analyses were performed to further explore potential sources of this heterogeneity. If heterogeneity cannot be resolved, narrative synthesis will be used in this study to analyze, summarize and compare the included studies. Funnel plots and Egger's test [15] were employed to assess publication bias. Symmetrical distribution of data points in the funnel plot and a p-value greater than 0.05 from Egger's test suggest no significant publication bias. In the event of evident publication bias, the trim-and-fill method will be employed to further assess the impact of publication bias on the results of the meta-analysis.

Results

Selection process

Figure 1 presents the PRISMA 2020 flowchart for the literature search and selection process. Our initial search retrieved a total of 8086 records from various databases (Web of Science, PubMed, MEDLINE, EMBASE, CINAHL, and the Cochrane Library) and manual searches ($n = 5$). After removing duplicates ($n = 1094$), 6992 records underwent title and abstract screening. Ultimately, 13 studies meeting the inclusion criteria were included in this review, encompassing a total of 13 prediction models. A table summarizing the number of retrieved records from each database is presented in Fig. 1.

Study characteristics

Thirteen studies were included in this review. Six employed retrospective cohort designs, five used prospective cohorts, and two were case-control studies (Table 1). Most studies ($n = 11$) utilized data from rehabilitation organizations, primarily hospitals. Public databases provided data for two studies. One study specifically focused on older adults with dementia (Table 1). The size of the study populations used to build the models ranged from 30 to 72,314 individuals (Table 1).

Table 2 summarizes the characteristics of the models used in the included studies. The studies employed various modeling techniques, including traditional logistic regression ($n = 4$), machine learning ($n = 4$), or a combination of both ($n = 5$). Only four studies incorporated external validation methods, while the remaining eight relied on internal validation (Table 2). Eight studies reported the model's discrimination performance, with AUC values ranging from 0.630 to 0.851 (Table 2). Two studies used calibration curves to assess calibration, while others reported metrics like sensitivity, specificity, positive

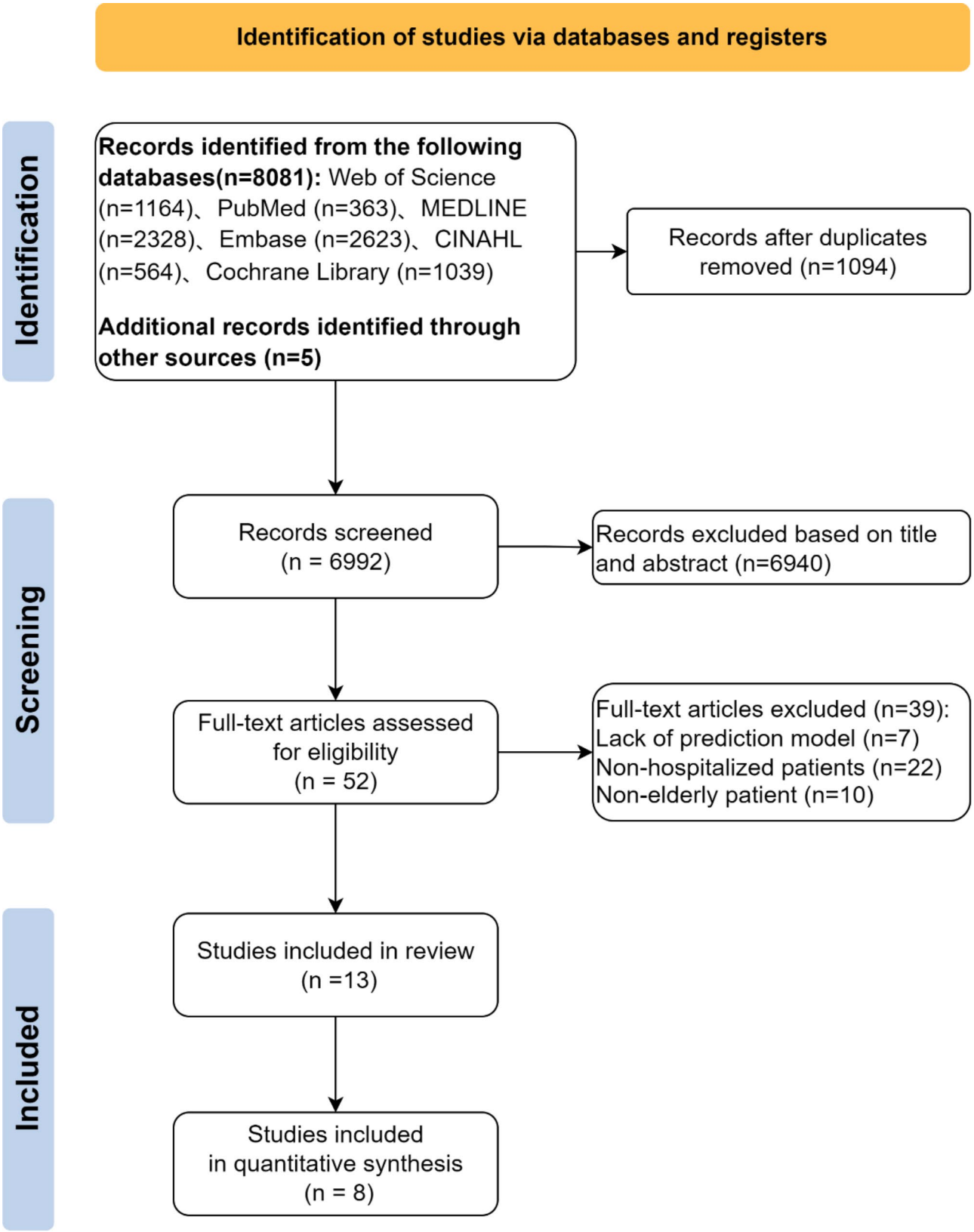

Fig. 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart of literature search and selection

Table 1 Overview of basic data of the included studies

Author (year)	Country	Study design	Data sources	Participants	Main outcome	Fall cases/ Sample size
Dormosh [16] 2023	Netherlands	Retrospective cohort study	An academic tertiary care hospital*	Inpatient(aged ≥ 70)	Fall(≥ 24 h of admission)	470/21,286
Adeli [17] 2023	Canada	Prospective cohort study	The Specialized Dementia Unit*	Hospitalized demented older inpatient	Falls while standing or walking	25/54
Zhao [18] 2020	China	case-control study	Three hospitals in Shanghai*	Inpatient(aged ≥ 65)	Injurious falls	115/345
Wijesinghe [19] 2020	Australia	Retrospective cohort study	MIMIC-III	Inpatient(aged ≥ 65)	Falls in Clinical Records	4314/12,911
Kawazoe [20] 2022	Japan	Retrospective cohort study	University of Tokyo Hospital*	Inpatient(aged ≥ 65)	Fall(> 48 h of admission and within 30 days)	1728/72,314
Chu [21] 2022	China	Retrospective cohort study	Taichung Veterans General Hospital*	Inpatient(aged > 65)	Fall(during the hospital stay)	349/1101
Alharbi [22] 2022	Saudi Arabia	Retrospective cohort study	SERV-112 and the SV-S2017 datasets	Older inpatient	Fall(during the hospital stay)	A:7295/9305 B:7408/9708
Peel [23] 2021	Australia	Prospective cohort study	Acute Care in Australian Hospitals*	Inpatient(aged ≥ 70)	Fall(during the hospital stay)	75/1288
Vratsistas-Curto [24] 2018	Australia	Prospective cohort study	General rehabilitation unit at a public hospital*	Older inpatient	fall(during rehabilitation stay)	41/300
Beauchet [25] 2018	France	Prospective cohort study	Acute care medical wards of Angers University Hospital*	Inpatient(aged ≥ 65)	Fall(during the hospital stay)	73/848
GholamHosseini [26] 2014	New Zealand	Prospective cohort study	North Shore Hospital*	Inpatient(aged ≥ 65)	Fall(during the hospital stay)	17/30
Neumann [27] 2013	Germany	Retrospective cohort study	An academic teaching hospital*	Inpatient(aged ≥ 65)	Fall(during the hospital stay)	508/4735
Marschollek [28] 2012	Germany	case-control study	Evangelisches Geriatriezentrum Berlin*	Older inpatient	Fall(during the hospital stay)	493/5176

MIMIC-III = Medical Information Mart for Intensive Care-III; A = Dataset SERV; B = SV-S2017 datasets; * Data obtained from routine electronic health record data

predictive rate, and negative predictive rate derived from the confusion matrix (Table 2). The final model predictors fell into four main categories: general demographics, physical and cognitive function, medications, and biochemical markers. The most frequently reported predictors (used in at least two studies each) were activity capacity ($n = 7$), history of falls ($n = 4$), medication ($n = 4$), mental cognition ($n = 4$), gender ($n = 2$), disease ($n = 2$), and vital signs ($n = 2$).

Risk of bias and applicability assessment

We used the PROBAST tool to evaluate the risk of bias and applicability of all 13 included models (shown in Fig. 2). A detailed quality assessment is provided in Appendix B.

Our analysis revealed a high risk of bias across all models. Eight studies had a high ROB due to unsuitable data sources (e.g., retrospective design). Similarly, eight studies were rated high ROB in the predictor domain due to the retrospective design lacking blinding, potentially influencing predictor assessment by outcome information. In the outcome domain, nine studies were judged high ROB given that they did not exclude outcome-related factors from the predictor definition, and one study was unclear due to missing information on the time interval between predictor assessment and outcome determination. Finally, all studies except Dormosh et al. [16] had high ROB in the analysis domain. Here, two studies fell short of the recommended sample size (EPV > 20), three studies involved the transformation of continuous data, and three studies excluded a portion of the data from the final analysis. Regarding data samples, two studies lacked data preprocessing (e.g., interpolation), and three used univariate analysis for predictor selection. Evaluation of model performance revealed that five studies omitted discrimination metrics, eleven omitted calibration metrics, and five neglected model fit assessment. Nine out of thirteen studies were classified as low risk for applicability, while four were considered high risk. All high-risk classifications stemmed from the participant domain. One study focused solely on older adults with dementia, a subgroup of the broader target population in this review. The remaining three high-risk

Table 2 Overview of the information of the included prediction models

Author (year)	Continuous variable processing method	Missing data handling	Modelling Methods	Validation method	Model performance			Model presentation	Final predictors	Interpretability
					Discrimination	Calibration	Others			
Dormosh [16] 2023	Continuous variable	Multiple imputation	LR + Lasso	10-fold cross-validation	AUROC: 0.695 (0.667–0.724)	Calibration curve	-	The formula of risk score	Fall history (0.34), Cardiac arrhythmias (0.35), Renal failure (0.33), Antipsychotics (0.48), Admission to neurologic department (0.53), Admission to emergency department (0.56), Heart rate (0.01), Katz ADL score (0.06), DOS score (0.04), Missing potassium (-0.38), Missing calcium (-0.25), Missing PaCO2 (-0.32), Missing DOS score (-0.41),	-
Adeli [17] 2023	Continuous variable	Imputation	ANN	Leave-one-out cross-validation	AUROC:0.762	-	Acc:0.731; Spe:0.732 Average Precision:0.499 Precision:0.441; Recall:0.728 F1:0.549	Prediction probability	-	NR
Zhao [18]2020	Continuous variable	NR	LR	Random split	AUROC A:0.874(0.784–0.964) B:0.847(0.771–0.924)	Calibration curve	Sen:0.692 Spe:0.855	Nomogram	History of fractures (1.67), Orthostatic hypotension (1.72), Functional status (1.07), Sedative-hypnotics (2.11), Level of serum albumin (2.53)	-
Wijesinghe [19] 2020	NR	NR	LR; SVM; RF	10-fold cross-validation	NR	NR	Precision:0.756; Recall:0.937; F1:0.836	Prediction probability	NR	NR
Kawazoe [20] 2022	Categorical variables	Multiple imputation	BERT + Bi-LSTM	Temporal validation	AUROC:0.851	NR	Sen:0.737; Spe:0.839; Precision:0.093; F1:0.165	Prediction probability	-	NR
Chu [21] 2022	Categorical variables	NR	DNN; XGBoost; LightGBM; RF; SGD; LR	Random split	AUROC:0.694	NR	Acc:0.730; Sens:0.694; Spe:0.694; Precision score:0.694; Recall score:0.694; F1:0.730	Prediction probability	-	PFI

Table 2 (continued)

Author (year)	Continuous variable processing method	Missing data handling	Modelling Methods	Validation method	Model performance			Model presentation	Final predictors	Interpretability
					Discrimination	Calibration	Others			
Alharbi [22] 2022	Categorical variables	NR	Catboost	Independent dataset validation	NR	NR	Dataset SERV: Acc:0.942, Sen:0.916, Spe:0.968, PPV:0.968,NPV:0.918 ,F1:0.941 Dataset SV: Acc:0.989, Sen:0.988, Spe:0.990, PPV:0.992, NPV:0.985, F1:0.990	Software-based platform	-	NR
Peel [23] 2021	Categorical variables	NR	LR	Independent dataset validation	AUROC: 0.700(0.630–0.760)	NR	Sen:0.72 spe:0.60	Scoring table	Sex (0.63), BMI (0.53), Fall in last 90 days (0.51), Balance problem (0.69), Psychological problems (0.86), Age (-0.03)	-
Vratsistas-Curto [24] 2018	Continuous variable	NR	LR	Bootstrap	AUROC: 0.730(0.660–0.810)	NR	-	Scoring table	Mobility/transfers, Mentalstatus/cognition, Male sex.	-
Beauchet [25] 2018	Categorical variables	Exclude	ANN	Random split	NR	NR	Acc:0.838; Sen:0.296; Spe:0.943 PPV:0.500; NPV:0.874; F1:0.372	Prediction probability	-	NR
Gholam-Hosseini [26] 2014	NR	NR	NR	Random split	NR	NR	Acc:0.740; Sen:0.850; PPV:0.850 F1:0.850	Scoring table	Real-time vital signs, Motion data, Medications, Falls history and muscle strength.	NR
Neumann [27] 2013	Continuous variable	-	LR; DT; Add-up model	Temporal validation	NR	NR	Sen:0.460; Spe:0.711; PPV:0.149 NPV:0.923	Scoring table	Mental alteration, Fall history, Insecure mobility	-
Marschollek [28] 2012	Continuous variable	Mean imputation	LR; DT	10-fold cross-validation	AUROC:0.63	NR	Acc:0.660; Sen:0.554; Spe:0.671 PPV:0.150; NPV:0.935; F1:0.237	Prediction probability	High age, Low Barthel index, Cognitive impairment, Multi-medication Co-morbidity	PFI

NR: Not report, Not applicable, A: Training data, B: test data, Multiple imputation: Imputation of missing values was performed multiple times by constructing estimation models, AUROC: Area Under the Receiver Operating Characteristic curve, Nomogram: A graphical interface that calculates the total score of the influencing factors to obtain the final prediction probability, LR: Logistic Regression, DT: Decision Tree, SVM: Support Vector Machine, RF: Random Forest, DNN: Deep neural network, XGBoost: eXtreme Gradient Boosting, LightGBM: Light Gradient Boosting Machine, SGD: Stochastic Gradient Descent, CatBoost: categorical boosting, ANN: Artificial Neural Network. Acc: Accuracy, Sen: Sensitivity, Spe: Specificity, PPV: Positive Predictive Value, NPV: Negative Predictive Value, F1: F1 Score (F1 Score =(2×Precision×Recall)/(Precision + Recall)), BERT: bidirectional encoder representations from transformers, Bi-LSTM: bidirectional long short-term memory, PFI: Permutation Feature Importance

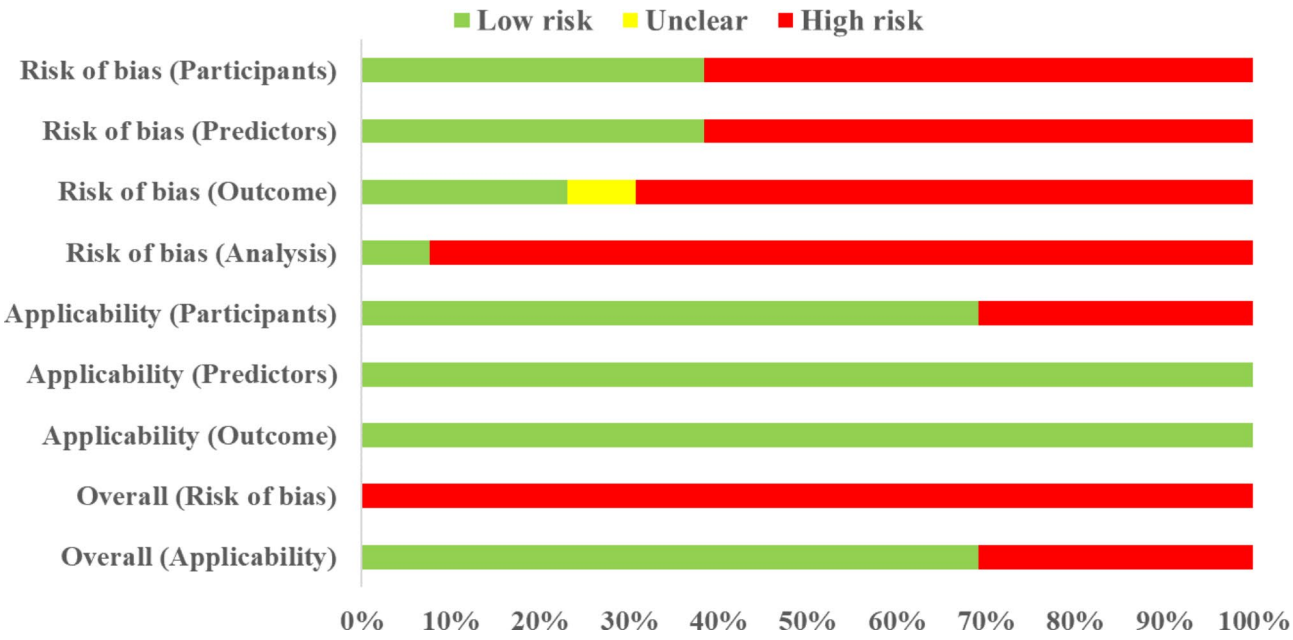

Fig. 2 Risk of bias and applicability assessment of included studies using the Prediction model Risk of Bias Assessment Tool (PROBAST)

studies did not define the age criteria for their older adult participants.

Meta-analysis of validation models included in the review

Due to the under-reporting of model assessment metrics, only eight studies were included in the meta-analysis for AUC. Notably, the prediction interval was significantly wider than the confidence interval, indicating substantial heterogeneity among the studies (shown in Fig. 3). Results of the sensitivity analysis (Appendix C) showed that after excluding individual studies in turn, the overall prediction interval was still significantly wider than the confidence interval, implying that there was still large heterogeneity. Subgroup analysis (Appendix D) revealed no significant difference in model performance between traditional logistic regression and machine learning algorithms. However, the within-group prediction interval was still significantly wider than the confidence interval, suggesting significant heterogeneity. Finally, Egger's test yielded a p-value of 0.102 indicating no significant publication bias.

Discussion

Hospital-acquired falls are serious adverse events, especially for older patients, leading to injuries, prolonged stays, and increased healthcare costs. Fall prevention is a crucial safety priority for healthcare providers, requiring individual fall risk assessments for each patient. This systematic review identified and assessed the quality of 13 studies on predictive models for falls in hospitalized older adults. The models exhibited significant performance variation in internal/external validation (AUROC: 0.630–0.851). However, the high risk of bias in all studies limits the real-world applicability of these findings.

This systematic review identified several critical methodological issues. Eight studies did not report how they handled missing data, while one study simply excluded it. This can introduce bias in effect size estimates and reduce the models' discriminative power. Multiple imputation [29] is the preferred approach for handling missing values in both model development and validation due to its accuracy and reduced bias. However, researchers should be mindful of "data leakage" [30]when using this method. Furthermore, four studies converted continuous variables into categorical ones. This can lead to information loss and reduced analytical power, ultimately resulting in lower model performance as documented in the literature [31].

Three of the included models used logistic regression, while the remaining five employed various machine learning algorithms. Machine learning is often viewed as superior to logistic regression for real-world data [32], which can be nonlinear and have complex relationships between features. This allows machine learning to handle large, high-dimensional datasets effectively. However, it should be borne in mind that machine learning models are not always superior [33]. In some cases, logistic regression models can be simpler and more effective. First, its simple form makes it easy to understand and interpret. Second, it can efficiently converge and provide stable results even with smaller datasets. The resulting regression coefficients indicate how strongly each variable influences the outcome. This interpretability is crucial for clinicians, as it allows them to identify key factors

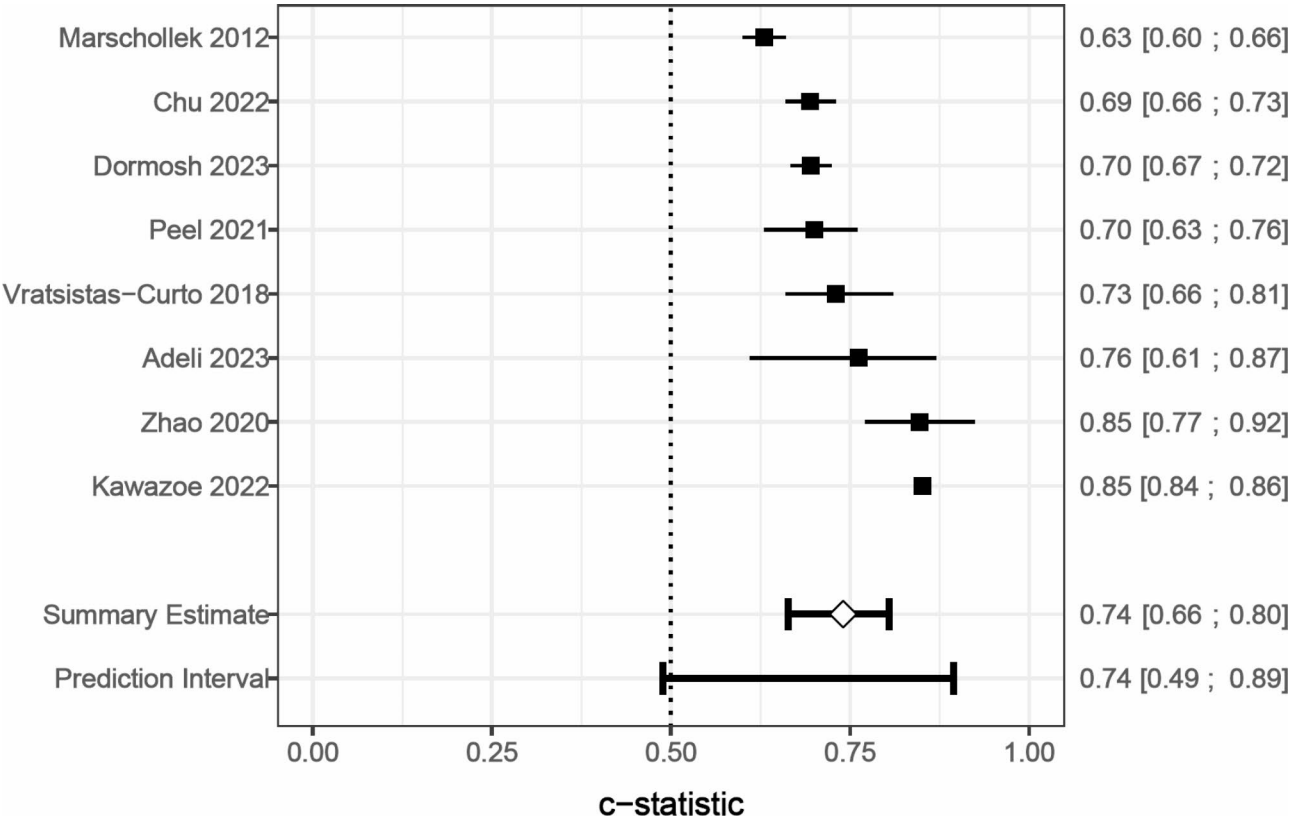

Fig. 3 Forest plots of fall risk prediction models for hospitalized older adults

in disease development and progression, informing preventive measures or treatment plans. Machine learning models are generally more complex than logistic regression, making them less interpretable, hence the "black box" label [34]. However, advancements are being made to enhance interpretability in these complex algorithms. SHapley Additive exPlanations (SHAP) is a popular example [35]. This game theory-based approach unveils the average contribution of each feature, enabling both global and local interpretability. Local interpretability allows clinicians to tailor rehabilitation programs to individual patients. Therefore, researchers must make trade-offs based on specific data characteristics when selecting modeling methods. To maximize the predictive performance and generalizability of the model, we recommend that researchers consider multiple modeling methods when constructing a prediction model.

Differing from static data, the construction of fall prediction models based on dynamically collected real-time or recent data holds broad prospects for development. In this study, the two studies that employed dynamic data to construct models both demonstrated favorable prediction accuracy (0.731–0.740). By segmenting the data or conducting time-series analysis to capture individual dynamic changes, it is possible to predict fall risk in real-time, which is crucial for the realization of early warnings. However, the data collection process may be plagued by issues of equipment stability and noise interference. The heterogeneity of the data further complicates data processing and increases the difficulty of model training. Consequently, it is imperative for the future to surmount the knowledge barriers between different fields through technological innovation and interdisciplinary collaboration.

Validation studies, both internal and external, can only assess a prediction model's performance in specific contexts, highlighting the need to confirm model robustness before clinical use [36]. In addition to conducting multicenter studies, researchers can utilize publicly available databases to enhance cost-effectiveness and generalizability by leveraging comprehensive data and larger datasets. However, it is crucial to attend to the temporal sequence between the extracted predictors and the occurrence of outcomes, neglecting this aspect could undermine the stability of the model and elevate the likelihood of missteps in clinical decision processes. Accurate reporting of model results is crucial for informed decision-making, transparency, and continuous model improvement. The PROBAST assessment tool emphasizes reporting on model discrimination (AUC ranges from 0.5 for random chance to 1 for perfect accuracy [37]) and calibration metrics. Additionally, clinical

applicability metrics like positive and negative predictive values can provide a more comprehensive assessment. In our study, although two studies reported calibration metrics, the provision of an Observed-to-Expected (O/E) ratio can offer more informative insights into the assessment of model calibration. For imbalanced datasets, the F1-Score and Matthews Correlation Coefficient (MCC) can be employed to comprehensively gauge model performance. Evaluating from multiple perspectives will provide a more holistic reflection of the predictive capabilities of the model, thereby ensuring the effectiveness of the chosen model in real-world applications. To improve reporting quality, researchers should strictly follow the Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [38].

Falls in older adults are a complex issue with both intrinsic (individual characteristics) and extrinsic (environmental) risk factors. Most falls involve a combination of advanced age, health conditions, and interactions with the environment [39]. Due to this complexity, predicting fall risk is challenging. This study summarized the most commonly reported influences on falls based on the final models' results. The top four factors identified were: mobility limitations, history of falls, medications, and mental cognition. While mobility testing is crucial for fall risk assessment, relying solely on a single test (e.g., Single Leg Stance Test, Timed Up and Go) is insufficient [40, 41]. Combining these tests with other factors improved the accuracy. Indeed, although numerous fall risk assessment tools exist, achieving both high sensitivity and specificity remains difficult [42]. Therefore, a more precise prediction model for hospitalized older adults is urgently needed for clinical application. A history of falls is a strong predictor of future falls and a major focus in clinical assessments [43]. This is likely due to both the physical consequences (functional decline) and psychological impact (fear of falling (FOF)) of falls. Notably, FOF is prevalent, affecting 40–73% of older adults with a history of falls, and even half of those without [44]. Polypharmacy and specific medications like cardiovascular and psychotropic drugs significantly increase fall risk in hospitalized older adults with high comorbidity [45]. Certain medications, including antiepileptic drugs, opioids, and those used in high quantities (polypharmacy), have been associated with an increased risk of falls in older adults. These factors should be considered during fall risk assessment. Cognitive impairment in older adults can impair their ability to cope with their environment, which can be detrimental to balance and gait [46]. However, more research is warranted to determine if there's a link between cognitive impairment and falling [47, 48]. In addition, falls can be influenced by various characteristics, including gender, medical conditions, and vital signs. Due to the complexity of falls, accurate assessment necessitates considering multiple factors. Future studies should prioritize incorporating well-established risk factors like those discussed above (mobility limitations, fall history, medications, and cognition) into fall risk models. Expanding the model's predictor base can address misclassification arising from variations in patient characteristics. However, it is important to avoid overfitting the model by introducing excessive complexity.

Strengths and limitations

Our study systematically reviewed multiple databases to evaluate research on fall prediction models for hospitalized older adults and conducted a critical assessment of the retrieved studies, providing comprehensive and objective evidence to support subsequent research. However, this study has several limitations that should be acknowledged. First, by only including English literature, we may have limited the diversity and generalizability of our findings. Additionally, although statistical tests indicated no significant publication bias, funnel plot and the exclusion of relevant studies from the grey literature databases may still lead to potential bias. Second, some studies lacked comprehensive reporting of results, hindering a meta-analysis on the calibration of the predictive models. Finally, the meta-analysis revealed a high degree of heterogeneity, which could be attributed to variations in study design, participant populations, and baseline fall risks. Although the current limitations preclude us from endorsing the clinical application of any specific model, our study can still provide valuable reference points for designing future high-quality studies with transparent reporting practices.

Implications

Our study aggregates and interprets the critical evidence related to fall risk factors in older adults admitted to hospitals, thereby serving as a cornerstone for the future development of precise and clinically actionable fall prediction models. Nevertheless, owing to the limitations in study design quality and the absence of robust model validation, the applied significance of the fall prediction models for hospitalized older adults as included in this research is not yet fully elucidated. The direction of future endeavors should be aimed at meticulous study design and the augmentation of external validation for established prediction models, with the objective of enhancing the broader applicability and generalizability of the research conclusions.

Conclusion

This study identified 13 studies with a total of 13 prediction models for fall risk in hospitalized older adults. The AUC values (0.630–0.851) indicate some discriminative

ability. However, all studies exhibited significant methodological shortcomings including a lack of rigorous experimental design or valid external validation. Consequently, we cannot recommend any model for clinical use at this stage. Future research should prioritize rigorous model validation adhering to the PROBAST standards for quality control. Additionally, leveraging big data for external validation can enhance model applicability and generalizability. Continuous optimization is crucial to maximize the model's practical value.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-025-05688-0.

Supplementary Material 1

Acknowledgements
Not applicable.

Author contributions
A.M. and H. Z. conceived the study; A.M. and J.S. designed the systematic review and conducted the literature search; A.M., J.S., M. R., and S. C. performed literature screening; A.M. and J.S. performed data extraction and risk of bias assessment; A.M. and J.S. wrote the manuscript, and H.Z. reviewed and revised it. All authors read and approved the final manuscript.

Funding
Not applicable.

Data availability
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All the studies included in this Systematic review were submitted to the ethical committee.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 19 April 2024 / Accepted: 7 January 2025
Published online: 14 January 2025

References

1. World Health Organization. Falls. 2021. https://www.who.int/news-room/fact-sheets/detail/falls. Accessed 26 April 2021.
2. Close JCT, Lord SR. Fall prevention in older people: past, present and future. Age Ageing. 2022;51:afac105. https://doi.org/10.1093/ageing/afac105.
3. Eglseer D, Hödl M, Lohrmann C. Six nursing care problems in hospitals: a cross-sectional study of quality of care. J Nurs Care Qual. 2019;34:E8–14. https://doi.org/10.1097/NCQ.0000000000000307.
4. Oliver D, Healey F, Haines TP. Preventing falls and fall-related injuries in hospitals. Clin Geriatr Med. 2010;26:645–92. https://doi.org/10.1016/j.cger.2010.06.005.
5. Matarese M, Ivziku D, Bartolozzi F, Piredda M, De Marinis MG. Systematic review of fall risk screening tools for older patients in acute hospitals. J Adv Nurs. 2015;71:1198–209. https://doi.org/10.1111/jan.12542.
6. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging Clin Exp Res. 2018;30:1–16. https://doi.org/10.1007/s40520-017-0749-0.
7. Solberg LM, Ingibjargardottir R, Wu Y, Lucero R. Nursing innovations in machine learning: using natural language processing in falls prediction. J Am Geriatr Soc. 2020;68:S48–9.
8. Zheng Q, Yang L, Zeng B, Li J, Guo K, Liang Y, et al. Artificial intelligence performance in detecting tumor metastasis from medical radiology imaging: a systematic review and meta-analysis. EClinicalMedicine. 2021;31:100669. https://doi.org/10.1016/j.eclinm.2020.100669.
9. Debray TPA, Damen JAAG, Snell KIE, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;i6460. https://doi.org/10.1136/bmj.i6460.
10. Moons KGM, De Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11:e1001744. https://doi.org/10.1371/journal.pmed.1001744.
11. 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. Ann Intern Med. 2019;170:51–8. https://doi.org/10.7326/M18-1376.
12. Cornell JE, Mulrow CD, Localio R, Stack CB, Meibohm AR, Guallar E, et al. Random-effects meta-analysis of inconsistent effects: a time for change. Ann Intern Med. 2014;160:267–70. https://doi.org/10.7326/M13-2886.
13. IntHout J, Ioannidis JP, Borm GF. The hartung-knapp-sidik-jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-laird method. BMC Med Res Methodol. 2014;14:25. https://doi.org/10.1186/1471-2288-14-25.
14. Higgins JPT. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60. https://doi.org/10.1136/bmj.327.7414.557.
15. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315:629–34. https://doi.org/10.1136/bmj.315.7109.629.
16. Dormosh N, Damoiseaux-Volman BA, van der Velde N, Medlock S, Romijn JA, Abu-Hanna A. Development and internal validation of a prediction model for falls using electronic health records in a hospital setting. J Am Med Dir Assoc. 2023;24:964–e9705. https://doi.org/10.1016/j.jamda.2023.03.006.
17. Adeli V, Korhani N, Sabo A, Mehdizadeh S, Mansfield A, Flint A, et al. Ambient monitoring of gait and machine learning models for dynamic and short-term falls risk assessment in people with dementia. IEEE J Biomed Health Inf. 2023;27:3599–609. https://doi.org/10.1109/JBHI.2023.3267039.
18. Zhao M, Li S, Xu Y, Su X, Jiang H. Developing a scoring model to predict the risk of injurious falls in elderly patients: a retrospective case–control study in multicenter acute hospitals. Clin Interv Aging. 2020;15:1767–78. https://doi.org/10.2147/CIA.S258171.
19. Wijesinghe YV, Xu Y, Li YF, Zhang Q. Pattern-based fall prediction using hospital clinical notes. IEEE/WIC/ACM Int Joint Conf Web Intell Intell Agent Technol (WI-IAT). 2020;433–7. https://doi.org/10.1109/WIIAT50758.2020.00064.
20. Kawazoe Y, Shimamoto K, Shibata D, Shinohara E, Kawaguchi H, Yamamoto T. Impact of a clinical text-based fall prediction model on preventing extended hospital stays for elderly inpatients: model development and performance evaluation. JMIR Med Inf. 2022;10:e37913. https://doi.org/10.2196/37913.
21. Chu WM, Kristiani E, Wang YC, Lin YR, Lin SY, Chan WC, et al. A model for predicting fall risks of hospitalized elderly in taiwan-a machine learning approach based on both electronic health records and comprehensive geriatric assessment. Front Med. 2022;9:937216. https://doi.org/10.3389/fmed.2022.937216.
22. Alharbi AH, Mahmoud H. Intelligent monitoring model for fall risks of hospitalized elderly patients. Healthcare. 2022;10:1896. https://doi.org/10.3390/healthcare10101896.
23. Peel NM, Jones LV, Berg K, Gray LC. Validation of a falls risk screening tool derived from InterRAI acute care assessment. J Patient Saf. 2021;17:E1152–6. https://doi.org/10.1097/PTS.0000000000000462.
24. Vratsistas-Curto A, Tiedemann A, Treacy D, Lord SR, Sherrington C. External validation of approaches to prediction of falls during hospital rehabilitation stays and development of a new simpler tool. J Rehabil Med. 2018;50:216–22. https://doi.org/10.2340/16501977-2290.
25. Beauchet O, Noublanche F, Simon R, Sekhon H, Chabot J, Levinoff EJ, et al. Falls risk prediction for older inpatients in acute care medical wards: is there an interest to combine an early nurse assessment and the artificial neural network analysis? J Nutr Health Aging. 2018;22:131–7. https://doi.org/10.1007/s12603-017-0950-z.

26. GholamHosseini H, Baig MM, Connolly MJ, Lindén M. A multifactorial falls risk prediction model for hospitalized older adults. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:3484–7. https://doi.org/10.1109/EMBC.2014.6944373.
27. Neumann L, Hoffmann VS, Golgert S, Hasford J, von Renteln-Kruse W. In-hospital fall-risk screening in 4,735 geriatric patients from the LUCAS project. J Nutr Health Aging. 2013;17:264–9. https://doi.org/10.1007/s12603-012-0390-8.
28. Marschollek M, Gövercin M, Rust S, Gietzelt M, Schulze M, Wolf K-H, et al. Mining geriatric assessment data for in-patient fall prediction models and high-risk subgroups. BMC Med Inf Decis Mak. 2012;12:19. https://doi.org/10.1186/1472-6947-12-19.
29. Nijman S, Leeuwenberg A, Beekers I, Verkouter I, Jacobs J, Bots M, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. J Clin Epidemiol. 2022;142:218–29. https://doi.org/10.1016/j.jclinepi.2021.11.023.
30. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. 2016;18:e323. https://doi.org/10.2196/jmir.5870.
31. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Stat Med. 2016;35:4124–35. https://doi.org/10.1002/sim.6986.
32. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit Care Med. 2016;44:368–74. https://doi.org/10.1097/CCM.0000000000001571.
33. Chen X, He L, Shi K, Wu Y, Lin S, Fang Y. Interpretable machine learning for fall prediction among older adults in China. Am J Prev Med. 2023;65:579–86. https://doi.org/10.1016/j.amepre.2023.04.006.
34. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). Comput Methods Programs Biomed. 2022;226:107161. https://doi.org/10.1016/j.cmpb.2022.107161.
35. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4768–77.
36. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ. 2016;i3140. https://doi.org/10.1136/bmj.i3140.
37. Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models: the C statistic. JAMA. 2015;314:1063–4. https://doi.org/10.1001/jama.2015.11082.
38. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162:55–63. https://doi.org/10.7326/M14-0697.
39. World Health Organization. WHO Global Report on Falls Prevention in Older Age. https://www.who.int/ageing/publications/Falls_prevention7March.pdf. Accessed 28 August 2020.
40. Beauchamp MK, Kuspinar A, Sohel N, Mayhew A, D'Amore C, Griffith LE, et al. Mobility screening for fall prediction in the Canadian longitudinal study on aging (CLSA): implications for fall prevention in the decade of healthy ageing. Age Ageing. 2022;51:afac095. https://doi.org/10.1093/ageing/afac095.
41. Tylman W, Kotas R, Kamiński M, Marciniak P, Woźniak S, Napieralski J, et al. Fully automatic fall risk assessment based on a fast mobility test. Sensors. 2021;21:1338. https://doi.org/10.3390/s21041338.
42. Kim YJ, Choi K, Cho SH, Kim SJ. Validity of the morse fall scale and the Johns Hopkins fall risk assessment tool for fall risk assessment in an acute care setting. J Clin Nurs. 2022;31:3584–94. https://doi.org/10.1111/jocn.16185.
43. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the updated American geriatrics society/british geriatrics society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc. 2011;59:148–57. https://doi.org/10.1111/j.1532-5415.2010.03234.x.
44. Murphy SL, Dubin JA, Gill TM. The development of fear of falling among community-living older women: predisposing factors and subsequent fall events. J Gerontol Biol Sci Med Sci. 2003;58:M943–7. https://doi.org/10.1093/gerona/58.10.M943.
45. Hartikainen S, Lonnroos E, Louhivuori K. Medication as a risk factor for falls: critical systematic review. J Gerontol Biol Sci Med Sci. 2007;62:1172–81. https://doi.org/10.1093/gerona/62.10.1172.
46. Taylor ME, Delbaere K, Lord SR, Mikolaizak AS, Brodaty H, Close JCT. Neuropsychological, physical, and functional mobility measures associated with falls in cognitively impaired older adults. J Gerontol Biol Sci Med Sci. 2014;69:987–95. https://doi.org/10.1093/gerona/glt166.
47. Lach HW, Harrison BE, Phongphanngam S. Falls and fall prevention in older adults with early-stage dementia: an integrative review. Res Gerontol Nurs. 2017;10:139–48. https://doi.org/10.3928/19404921-20160908-01.
48. Efendioglu EM, Cigiloglu A, Ozturk ZA. The role of comprehensive geriatric assessment in predicting fall risk. Ir J Med Sci. 2023;192:303–10. https://doi.org/10.1007/s11845-022-02978-z.

Publisher's note