Fall Prediction in Older Adults: A Model Based on Fall-Trajectory Predictors Collected in Patients' Homes

Amadou M. Djiogomaye Ndiaye
Laboratoire Vie-Santé, UR 24 134, Faculté de Médecine
Chaire d'excellence chez Fondation Partenariale de l'Université de Limoges I Institut Omega Health
Limoges, France
e-mail: amadou_maguette_djio.ndiaye@unilim.fr

Laurent Billonnet
XLIM-SRI - Systèmes et Réseaux Intelligents (444315) – XLIM
Limoges, France
e-mail: laurent.billonnet@unilim.fr

Michel Harel
Laboratoire Vie-Santé, UR 24 134, Faculté de Médecine
INSPÉ de Limoges, Université de Limoges
Limoges, France
Institut de Mathématiques de Toulouse, UMR CNRS 5 219, Université Paul Sabatier
Toulouse, France
e-mail: michel.harel@unilim.fr

Achille Tchalla
Laboratoire Vie-Santé, UR 24 134, Faculté de Médecine
Service de Médecine Gériatrique, CHU de Limoges Dupuytren
Pôle Hospitalo-Universitaire de Gérontologie clinique, Service de Médecine Gériatrique, Unité de Prévention de Suivi et d'Analyse du Vieillissement, CHU de Limoges Dupuytren
Chaire d'excellence chez Fondation Partenariale de l'Université de Limoges I Institut Omega Health
Limoges, France
e-mail: achille.tchalla@unilim.fr

***Abstract*—Falls among older adults are a major public health concern due to their frequency, consequences and impact on autonomy and mortality. The Risk Of Falling (ROF) is linked to three dimensions: physical/organic, socio-environmental and thymic/cognitive. Identifying individuals at high risk is essential to implementing personalized prevention strategies. While fall history is a well-known predictor, the integration of multi-dimensional health data and interpretable machine learning models may enhance prediction accuracy. We conducted a retrospective analysis of 1,648 older adults who underwent a Comprehensive Geriatric Assessment (CGA) at two time points. Based on clinical, functional, cognitive and psychosocial variables, we developed and compared four supervised classification models: logistic regression, Support Vector Machine (SVM), random forest and eXtreme Gradient Boosting (XGBoost). Predictive performance was evaluated using Area Under the receiver operating characteristic Curve (AUC), F1-score and Brier score. SHapley Additive exPlanations (SHAP) values were used to interpret variable contributions at the individual level. XGBoost and random forest models demonstrated the best performance (AUC = 0.76 and 0.77, F1-score = 0.72 and 0.73, Brier score = 0.19 for both). SHAP analysis confirmed that fall history was a strong predictor but not the sole contributor to the model's decisions. Functional limitations, low Activities of Daily Living (ADL) and low Instrumental Activities of Daily Living (IADL), impaired physical performance (low Short Physical Performance Battery (SPPB)), pathological Single Leg Balance (SLB) and cognitive scores (Mini-Mental State Examination (MMSE)) also played substantial roles. Misclassified cases illustrated the importance of multidimensional balance in the model's outputs. Our findings support the use of interpretable machine learning models, particularly XGBoost, for personalized fall risk prediction in older adults. Beyond fall history, a combination of physical, cognitive and psychosocial variables contributes meaningfully to risk estimation. Such models may help guide targeted preventive interventions in geriatric practice, provided operational complexity is managed to allow real-world clinical integration.**

Keywords-fall; older population; prevention; personalized medicine; AI.

I. INTRODUCTION

This article is an extended version of the international conference paper entitled *"Enhancing Fall Prediction in Older Adults: A Data-Driven Approach to Key Parameter Selection"* [1]. In this extended version, some models have been upgraded by including dyslipidemia, a cardiovascular factor, among the predictive variables for falls. However, we retain XGBoost as our final model, since it remains one of the most effective approaches for ensuring both high predictive performance and interpretability in personalized prediction.

According to the World Health Organization (WHO), older individuals are those aged ≥ 60 years. The proportion of older individuals worldwide is expected to nearly double between 2015 and 2050, increasing from 12% to 22% [2]. The National Institute of Statistics and Economic Studies (INSEE) estimates that one in three individuals in France will be aged ≥ 60 years by 2060, compared to one in four individuals in 2021 [3]. Aging leads to a gradual decline in functional capacity, increasing the ROF [4]. Falls in older adults represent a major public health concern due to their high frequency, their functional, psychological and economic consequences, as well as their impact on mortality. In the study by Tan et al. [5], falling was identified as one of the main predictive factors integrated into a model designed to identify long-term care patients at highest risk of death. Similarly, Shaik et al. [6] highlighted that, in both older and younger individuals, falls, along with bone pathologies, are among the primary causes of hip fractures.

Fall prevention has always been a central focus in medical practice, notably through clinical test batteries or by adjusting specific functions according to identified predictive factors, generally using linear regression models (LRMs), after grouping patients based on shared health characteristics. While traditional regression models have long been the standard tool for analyzing risk factors, machine learning methods now offer improved predictive performance by accounting for complex interactions between variables.

We developed predictive models using as input data the factors identified in various fall trajectories. The objective is to evaluate whether these variables are sufficiently discriminative to power an effective predictive model, among all those tested and thereby contribute to a targeted and personalized fall risk prevention strategy. Early identification of ROF facilitates the administration of personalized interventions for individuals [7].

Most recent studies predict falls using sensors or Electronic Health Records (EHRs). With data collected directly from elderly individuals' homes, our objective is to develop an effective predictive model using the fewest possible features.

In this study, we evaluated and compared several classification algorithms to predict fall risk based on clinical, functional and psychosocial data collected from a CGA. Model interpretability was ensured using SHAP values, in order to facilitate clinical understanding of the results and to precisely identify the factors that most contributed to the prediction of fall risk.

II. Materials and Methods

A. Study Design

Our study is based on a dataset collected between September 2011 and September 2023 through multiple home visits conducted by the Unit for Prevention, Monitoring and Analysis of Ageing (UPSAV – *Unité de Prévention, de Suivi et d'Analyse du Vieillissement*) at Limoges University Hospital, Limoges, France. The UPSAV team comprises nurses, geriatricians and other healthcare professionals. Each patient underwent an initial visit, followed by a second visit six months later and a third visit one year after the second. If the patient remains in the study after the third visit, subsequent visits occur annually. The study includes men and women aged 60 and older. To be eligible, participants had to meet the following criteria:

- Provide written informed consent, either personally or through a legal representative.
- Not be enrolled in a clinical trial that modifies their standard medical management.
- Not have progressive pathologies that could significantly affect short-term prognosis.
- Not reside in a long-term care unit or a nursing home.
- Be covered by social security at 100%.

B. Falls and Comprehensive Geriatric Assessment

During the Follow-up, a fall was defined as unintentionally coming to rest on the ground or other lower level not as a result of a major intrinsic event (e.g., myocardial infarction, stroke, or seizure) or an overwhelming external hazard (e.g., hit by a vehicle) [8], [9]. Each patient underwent a CGA and received a personalized care plan. The CGA is a multidimensional and standardized approach designed to enhance clinical practices in the care of older adults through a comprehensive health assessment. CGA are widely used to evaluate the physical, cognitive, social and medical factors associated with fall risk in older adults [10]. Although they provide valuable clinical information, CGAs often involve numerous variables and can be time-consuming to administer and interpret, particularly in home care settings. This highlights the growing need for efficient and scalable tools that can help prevent falls without increasing the burden on caregivers or patients.

Falls may occur repeatedly within a year. In geriatric practice, individuals who experience at least two falls within a 12-month period are classified as "fallers" [11].

A holistic fall prediction approach considers three key dimensions:

- The physical/organic dimension gathers data related to an individual's medical history and current symptoms, diagnosis of underlying health issues and treatment effectiveness.
- The thymic/cognitive dimension refers to an individual's mental, emotional and cognitive states.
- The socio-environmental dimension refers to age, gender, family and social support, housing conditions, home configuration, the presence of slippery rugs, stairs without railings, uneven surfaces and inadequate lighting.

Evaluating the ROF involves at least a gait and balance assessment of the physical/organic dimension and the age and gender of the socio-environmental dimension. Data involving the thymic/cognitive dimension allow for a comprehensive review of the potential causes of a fall. The term "dimension" refers to the types of factors that contribute to the ROF and their evaluation.

Hospitalized patients often receive incomplete health assessments across all dimensions. Our home-collected data encompass features from all three dimensions.

C. Data Collection and Variable Processing

Covariates included fall occurrences, cardiovascular risk factors, socio-environmental characteristics and the CGA summary. Fall occurrences refer to falls that occurred between visits.

Socio-environmental characteristics assessed in the home included gender, age, lifestyle, housing conditions, presence of an elevator, long-term illness status, leisure activities, social activity, human assistance and pet ownership.

Cardiovascular risk factors considered were hypertension, diabetes, dyslipidemia, obesity and tobacco use.

The CGA summary encompassed multiple functional and cognitive assessments, including:

- Verbal fluency test [12],
- Single Leg Balance (SLB) test, scored 0-60 seconds [13],
- Clock-drawing test (CDT), scored 0-5 [14],
- Instrumental Activities of Daily Living (IADL), scored 0-8 [15],
- Mini-Mental State Examination (MMSE), scored 0-30 [16],
- Mini Nutritional Assessment (MNA), scored 0-30 [17],
- Short Physical Performance Battery (SPPB), scored 0-12 [18],
- Geriatric Depression Scale (GDS), scored 0-30 [19].

For consistency, in the rest of the document, we added 'Pathological' to the feature names SLB test, CDT, Verbal Fluency and GDS to indicate whether the test result is positive or not.

D. Data analysis

In our study, the sample size decreased from 1,648 patients at the first visit to 954 patients followed up at the second visit. A descriptive analysis was conducted to provide an overview of the study variables and their distribution between individuals who had fallen and those who had not. Pearson's Chi-squared test was used for categorical variables, while the Wilcoxon rank-sum test was applied to continuous variables. The significance threshold for all statistical tests was set at a p-value (P) < 0.05 and all reported P-values were two-tailed. The p-value or probability value is a statistical measure ranging between 0 and 1. It expresses the probability of obtaining a result at least as extreme as the one observed under the assumption that the null hypothesis (H_0) is true. The null hypothesis used as the starting point of a statistical test states that there is no effect, no difference, or no relationship between the variables under study. According to the most commonly accepted convention a result is considered statistically significant when $p < 0.05$. In this case, the probability of obtaining the observed data (or more extreme outcomes) under H_0, is less than 5%. The null hypothesis is therefore rejected in favor of the alternative hypothesis (H_1), suggesting the existence of an effect or a difference. All statistical analyses were performed using R software (version 4.4.0, R Foundation for Statistical Computing, Vienna, Austria).

E. Model Development Using Supervised Machine Learning

The construction of a predictive model relies primarily on selecting a limited number of relevant variables. In geriatrics, preventive strategies implemented by geriatricians traditionally rely on "predictive factors" identified using logistic regression models (LRM). These factors correspond to variables significantly associated with fall risk across different patient groups (or clusters), formed based on longitudinal (or panel) data collected at multiple time points during the study.

In our work, after identifying the fall trajectories specific to the study population, we extracted the most explanatory variables for each of these trajectories. These predictive variables then served as the basis for building several predictive models, which we compared in order to evaluate their performance.

We developed a fall risk prediction model by selecting the best-performing algorithm among four classifiers: logistic regression, Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost) and Random Forest. Logistic regression is a linear supervised classification model particularly suited for binary problems . SVM, on the other hand, aims to maximize the margin between classes using an optimal hyperplane [20]. Ensemble models such as XGBoost and Random Forest rely on aggregating multiple decision trees: the former through a sequential boosting process and the latter through a bagging mechanism, both of which enhance model accuracy and robustness [21], [22].

To optimize the performance of each classifier, we used the RandomizedSearchCV method, which randomly explores a subset of hyperparameter combinations within a defined search space. Unlike GridSearchCV, which exhaustively evaluates all possible combinations, this approach reduces computational cost while efficiently exploring influential parameters through cross-validation. Finally, to calibrate the predicted probabilities of the models, we applied 5-fold cross-validation calibration using CalibratedClassifierCV (with cv=5) before evaluating final performance on the test set.

No missing data were observed among the variables included in the analysis. To address class imbalance, the RandomUnderSampler method was applied, which consists of randomly removing observations from the majority class to rebalance the dataset. Given the sensitive and real nature of health data, no synthetic oversampling method was used. The dataset was randomly split into a training set (70%) and a test set (30%).

Model performance was evaluated on both the training and test sets using several metrics: Area Under the Curve (AUC), accuracy, precision, recall, specificity, F1-score and Brier score [23], [24], [25]. Among these, AUC, F1-score and Brier score were selected as the main evaluation indicators. AUC assesses the model's discrimination ability, the F1-score captures the balance between precision and recall, while the Brier score measures the accuracy of probabilistic predictions; it is calculated as the mean squared difference between predicted probabilities and actual outcomes. A high AUC and F1-score, combined with a low Brier score, indicate good classification performance and accurate probability estimation.

For model interpretation, SHAP values were computed to quantify the contribution of each variable to the individual prediction of fall risk. SHAP is an explainable AI method that provides insights into the contribution of each feature both globally (across the entire dataset) and locally (for individual predictions) [26]. All algorithms were implemented in Python 3.10.16 (Python Software Foundation, Wilmington, DE). Variable preprocessing was performed using OneHotEncoder for categorical variables and StandardScaler for numerical variables, via the scikit-learn library.

III. RESULTS

A total of 1,648 individuals met the inclusion criteria for the study. Table I presents the baseline socio-environmental and health characteristics of the sample that significantly differentiate fallers from non-fallers. Among the older adults included, 1,113 (68%) were women and 535 (32%) were men. Additionally, 73% had hypertension and only 288 participants (17%) engaged in social activities. The mean age of participants was 83 ± 6 years. Regarding falls, 823 participants (approximately 50%) had experienced a fall during the previous year. Concerning housing conditions, 991 (60%) were homeowners. Furthermore, 449 participants (27%) were classified as having depression.

TABLE I. OVERVIEW OF BASELINE CHARACTERISTICS ACCORDING TO FALLS OF THE STUDY

Features of the study	Total sample (N = 1,648) n (%)	Falls of the study		p-value*
		No falls (n = 794, 48.2%)	Falls (n = 854, 51.8%)	
Woman	**1,113 (68%)**	**500 (63%)**	**613 (72%)**	**<0.001**
Age, m ± SD, years	**83 ± 6**	**82 ± 6**	**83 ± 6**	**0.001**
Diabetes	**339 (21%)**	**146 (18%)**	**193 (23%)**	**0.035**
Leisure	**1,377 (84%)**	**689 (87%)**	**688 (81%)**	**<0.001**
Social activity	**288 (17%)**	**162 (20%)**	**126 (15%)**	**0.003**
Human assistance	**1,402 (85%)**	**644 (81%)**	**758 (89%)**	**<0.001**
ADL, m ± SD	**5 ± 1**	**5 ± 1**	**5 ± 1**	**<0.001**
IADL, m ± SD	**6 ± 2**	**6 ± 2**	**5 ± 2**	**<0.001**
MMSE, m ± SD	**23 ± 7**	**24 ± 7**	**23 ± 7**	**0.006**
Pathological CDT	**585 (35%)**	**244 (31%)**	**341 (40%)**	**<0.001**
Pathological verbal fluency	**672 (41%)**	**269 (34%)**	**403 (47%)**	**<0.001**
MNA, m ± SD	**24 ± 4**	**24 ± 4**	**23 ± 4**	**<0.001**
SPPB, m ± SD	**7 ± 4**	**7 ± 4**	**6 ± 4**	**<0.001**
Pathological GDS	**449 (27%)**	**176 (22%)**	**273 (32%)**	**<0.001**
Pathological SLB	**708 (43%)**	**261 (33%)**	**447 (52%)**	**<0.001**

*Pearson's Chi-squared test; Wilcoxon rank sum test. Statistically significance (p-value < .05). m, mean; SD, Standard deviation; SLB, Single leg balance; CDT, Clock-drawing test; ADL, Activities of Daily Living; IADL, Instrumental Activities of Daily Living; MMSE, Mini-Mental State Examination; MNA, Mini Nutritional Assessment; SPPB, Short Physical Performance Battery; GDS, Geriatric Depression Scale. Data are shown as the number (percentage) or mean ± SD unless otherwise indicated.

In Table II, which presents the variables included in our predictive models, it is observed that among the 954 participants included in the study, 48.6% reported at least one fall prior to the follow-up period. Fallers exhibited several characteristics that were significantly different ($p \leq 0.05$) from non-fallers. Fallers were predominantly women (74% vs 66%). Their functional and physical abilities were generally more impaired: lower ADL scores, reduced SPPB scores (6 ± 4 vs 8 ± 3) and lower IADL scores. Depression, as indicated by a pathological GDS score, was more frequent among fallers (31% vs 20%) and postural instability, assessed by a pathological one-leg stance test, was observed in 46% of fallers compared to 34% of non-fallers. Participation in leisure activities was also slightly lower among fallers (86% vs 91%), which could reflect behavioral withdrawal or functional restriction.

TABLE II. OVERVIEW OF THE SIX-MONTH INPUT FEATURES USED IN OUR PREDICTIVE MODELS

Features	Falls of the study (N = 954)			p-value*
	Total sample (N = 954) n (%)	No falls (n = 490, 51.4%)	Falls (n = 464, 48.6%)	
Woman	**664 (70%)**	**321 (66%)**	**343 (74%)**	**0.005**
Hypertension	688 (72%)	353 (72%)	335 (72%)	0.96
Dyslipidemia	453 (47%)	237 (48%)	216 (47%)	0.57
Obesity	254 (27%)	122 (25%)	132 (28%)	0.21
Leisure	**843 (88%)**	**445 (91%)**	**398 (86%)**	**0.015**
MMSE, m ± SD	25 ± 6	25 ± 6	25 ± 6	0.13
SPPB, m ± SD	**7 ± 4**	**8 ± 3**	**6 ± 4**	**<0.001**
ADL, m ± SD	**5 ± 1**	**6 ± 1**	**5 ± 1**	**<0.001**
IADL, m ± SD	**6 ± 2**	**7 ± 2**	**6 ± 2**	**<0.001**
Pathological GDS	**238 (25%)**	**96 (20%)**	**142 (31%)**	**<0.001**
Pathological SLB	**381 (40%)**	**167 (34%)**	**214 (46%)**	**<0.001**

*Pearson's Chi-squared test; Wilcoxon rank sum test. Statistically significance (p-value < .05).m,mean; SD, Standard deviation; SLB, Single leg balance; CDT, Clock-drawing test; ADL, Activities of Daily Living; IADL, Instrumental Activities of Daily Living; MMSE, Mini-Mental State Examination; MNA, Mini Nutritional Assessment; SPPB, Short Physical Performance Battery; GDS, Geriatric Depression Scale. Data are shown as the number (percentage) or mean ± SD unless otherwise indicated.

In contrast, hypertension, dyslipidemia, obesity and MMSE scores were not statistically associated with falls in this cohort.

These results support the hypothesis of a multifactorial etiology of falls, primarily driven by physical function impairment, loss of autonomy, mood disorders, depression and postural balance issues.

A comparison of Table I and Table II shows that variables such as hypertension, obesity and dyslipidemia are predictive factors of falls but do not significantly differentiate fallers from non-fallers. The remaining variables reported in Table II are also significant in Table I.

Fig. 1 presents the AUC of the four models evaluated for predicting fall risk, namely logistic regression, SVM, XGBoost and random forest.

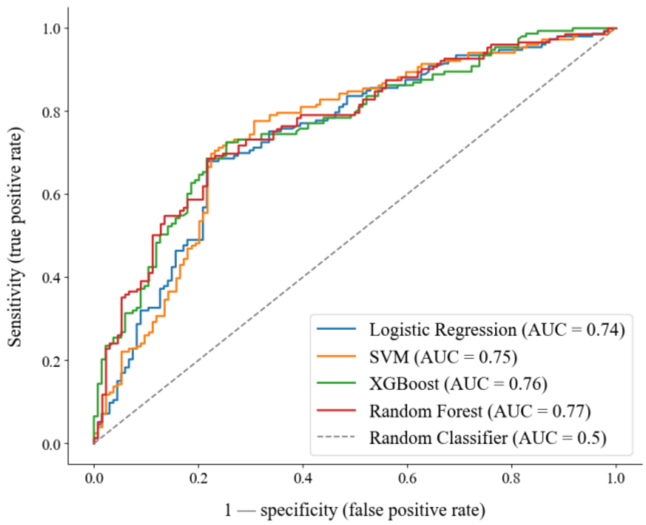

Figure 1. Area Under the Curve (AUC) of the Different Models

Table III reports the performance metrics of the different models. All models achieved an identical precision of 0.78, with balanced F1-scores ranging between 0.71 and 0.73, indicating comparable overall classification performance.

TABLE III. SUMMARY OF PREDICTIVE PERFORMANCE OF THE DIFFERENT MODELS

Metrics	Logistic Regression	SVM	XGBoost	Random Forest
AUC	0.74	0.75	**0.76**	**0.77**
Accuracy	0.73	0.71	0.72	0.73
Precision	0.78	0.78	0.78	0.78
Recall	0.68	0.65	0.67	0.68
Specificity	0.78	0.78	0.78	0.78
F1 score	0.73	0.71	**0.72**	**0.73**
Brier score	0.20	0.20	**0.19**	**0.19**

However, XGBoost and Random Forest show better areas under the ROC curve, with AUC values of 0.76 and 0.77 respectively (see Fig. 1), suggesting higher discriminative ability compared to logistic regression (AUC = 0.74) or SVM (AUC = 0.75). Recall is slightly lower for XGBoost (0.67) than for Random Forest (0.68), which may reflect a tendency to under-detect certain fall cases. Finally, the lowest Brier scores (0.19) are achieved by XGBoost and Random Forest, indicating better probabilistic calibration of predictions. Thus, although all models perform similarly in classification, Random Forest appears to offer the best trade-off between discrimination and calibration.

XGBoost and Random Forest are the models with the best overall performance. Both are tree-based methods; while Random Forest makes binary decisions, XGBoost has the advantage of computing individualized probabilities, which makes it more suitable for personalized care approaches. To better understand the contribution of each variable to the model's predictions, we apply SHAP to XGBoost.

The analysis of SHAP values presented in Fig. 2 highlights both the relative importance and the direction of effect of each variable in predicting fall risk within the XGBoost model.

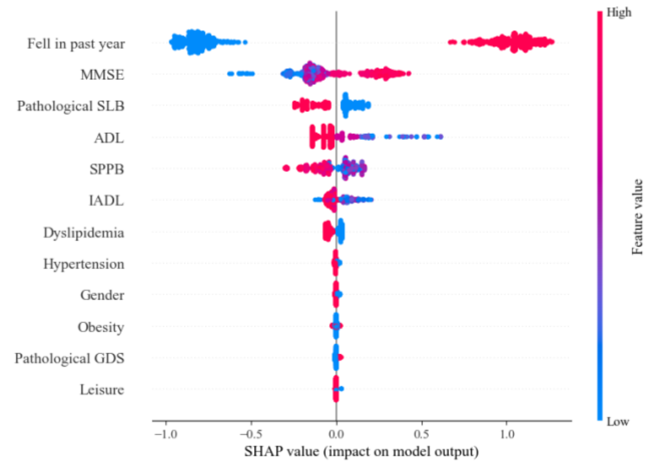

Figure 2. Impact of the Different Variables on the Best Model (XGBoost)

The use of SHAP values provides transparent model interpretation and may help inform priorities for targeted preventive strategies. A low score (values in blue) contributes significantly to risk reduction, whereas a high score (shown in red) is associated with increased predicted risk.

Among all the variables considered, fall history emerges as the most influential factor thereby confirming the strong predictive power of prior fall events. Physical performance, as assessed by the SPPB score also plays a central role in fall-risk prediction, low SPPB values (indicating physical impairment) are strongly associated with higher risk. Pathological single-leg stance reflecting balance impairments does not appear to be correlated with elevated fall risk ; in some cases, it may even be linked to severely limited mobility thereby reducing exposure to risk through restricted movement.

At the cognitive level, the MMSE score shows a more nuanced relationship while low scores are generally considered a risk factor, their impact appears less pronounced in the model. Conversely, higher scores may counterintuitively be associated with increased risk possibly due to overconfidence or engagement in unsafe physical activities.

The ADL and IADL scores indicators of functional autonomy exhibit patterns consistent with clinical evidence reduced functional capacity is generally associated with increased fall risk. However, very low IADL scores may not strongly correlate with higher risk suggesting that advanced dependency could reduce exposure to hazardous situations.

Dyslipidemia reflecting cardiovascular impairment is unexpectedly associated with a lower risk of falls potentially indicating a tendency to avoid physical activity due to fear of falling.

Other variables including hypertension, obesity, gender, and the presence of depression (pathological GDS), exert a more moderate or marginal influence on model predictions. Participation in leisure activities shows a modest protective effect, although its overall contribution to fall-risk prediction remains limited.

In summary, this analysis underscores that the most influential predictors of fall risk are functional and physical domains, while cognitive and psychosocial dimensions exert secondary effects.

After examining the impact of each variable on the model's predictions, we now turn to some examples of personalized predictions.

The personalized predictions will be evaluated using the final selected XGBoost model. XGBoost is a gradient boosting ensemble algorithm that aggregates multiple weak decision trees to produce a high-performing predictive model [22]. In binary classification, it generates a raw output in log-odds, which is then transformed by the logistic function to obtain a probability. The log-odds (logarithm of the odds) is a way to transform a probability into a value that can range from –∞ to +∞. The raw output value of XGBoost is the weighted sum of the decision trees:

$$f(x) = \sum_{k=1}^{K} T_k(x)$$

where:

- $T_k(x)$ is the output of the k-th tree for the observation,
- K is the total number of trees,
- $f(x)$ is the raw model output, expressed in log-odds.

We then transform the raw output $f(x)$ into a probability $p(x)$ with the sigmoid function :

$$p(x) = \sigma(f(x)) = \frac{1}{1 + e^{-f(x)}}$$

where the sigmoid function is defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

The reference value (base value) is the mean of $f(x)$ and the associated probability is the overall prevalence in the training sample.

Fig. 3 below illustrates a correctly predicted low fall risk ($f(x)$ = 0.23) compared with the base value of 0.48. Protective factors such as a lower MMSE score of 18, preserved ADL of 6, and a high IADL of 8 strongly contributed to reducing the predicted risk. A history of falls also contributed to lowering the prediction. Although risk-increasing variables such as the absence of a pathological one-leg stance and a low SPPB score of 4 were present, they were outweighed by the protective factors.

Fig. 4 below illustrates a case classified as high fall risk, with a predicted probability of 0.65. However, the prediction is incorrect; in the collected data, the patient did not fall. Several strong risk factors were present including a history of falls, dyslipidemia, low IADL (6) and a low SPPB score (9) all of which contributed to increasing the predicted risk. Nevertheless, these were insufficiently weighted by the model while mitigating factors such as a relatively high MMSE score (23), a non-pathological SLB and a moderate ADL score (6) overly influenced the output leading to a misclassification. This highlights the model's limitation in edge cases where compensatory features may mask critical risks.

These personalized predictions of two different patients highlight that the model's outputs do not depend solely on fall history, even though it is the strongest predictor among all variables (Fig. 2). The trends observed in the SHAP values of all variables in Fig. 2 are confirmed by the prediction shown in Fig. 3.

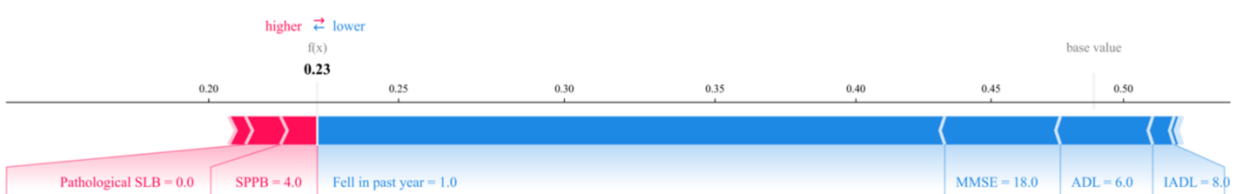

Figure 3. Correct Prediction with SHAP

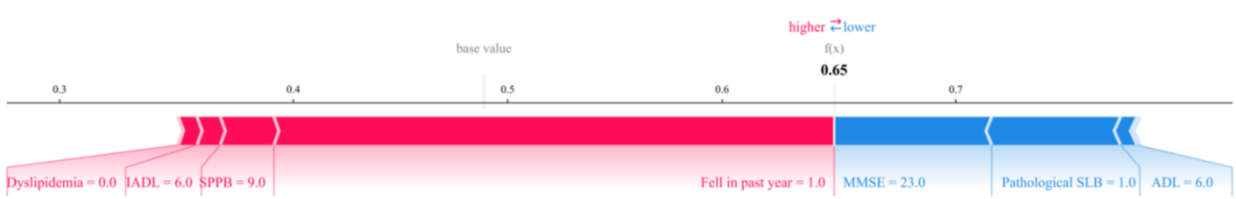

Figure 4. Incorrect Prediction with SHAP

IV. DISCUSSION

The evaluation of various fall risk predictors (see Table II), based on data from patients who completed both the first and second visits, revealed that most variables showed significant differences between fallers and non-fallers. This highlights the importance of identifying predictive factors within the least stable clusters (i.e., those in which falls were observed), as opposed to more stable clusters. Among the variables analyzed, the following were significantly different depending on group membership (fallers vs. non-fallers): gender (female), ADL score, IADL score, SPPB score, presence of a pathological GDS score, pathological SLB and participation in leisure activities.

Among these variables, only sex and participation in leisure activities pertain to the socio-environmental domain and could be collected in other protocols. The remaining variables are scores derived from the CGA conducted at the patients' homes. These findings support the hypothesis that a holistic approach is necessary for predicting fall risk. Specifically, the pathological GDS score reflects the thymic/cognitive dimension, while the ADL, IADL and SPPB scores, along with the pathological one-leg stance, reflect the physical/organic dimension.

Using the variables most significantly associated with fall risk (see Table II) as input data represents a relevant strategy, as the model's objective is to differentiate fallers from non-fallers in a personalized manner. In order to remain aligned with the clinical approach of identifying predictive factors to develop targeted prevention plans, all variables identified (see Table II) were retained for model training. Fig. 2 confirms the importance of these variables, showing that they rank among the most influential in the XGBoost model, with the exception of gender and pathological GDS score, which were replaced by dyslipidemia and MMSE score in terms of predictive weight. The integration of dyslipidemia, a cardiovascular risk factor and the MMSE score, a marker of cognitive function, further reinforces the model's holistic approach.

Not every feature within the three ROF dimensions is a predictive factor for falls. The effectiveness of a predictive factor depends on its statistical significance, correlation with fall occurrences and its interaction with other variables across the physical/organic, socio-environmental and cognitive dimensions. In some studies, the identified predictive variables did not encompass all three dimensions of ROF. Kawazoe et al. [27], Ikeda et al. [28] and Cella et al. [29] demonstrated that age category related to socio-environmental was a predictor of falls, suggesting a strong association between age and falls. Bath et al. [30] found that the predictive variables related to the socio-environmental dimension are diverse and varied, contributing to effective prevention. In fact, a higher number of variables related to gait and balance is associated with a more robust predictive model for falls.

In the literature review conducted by Rubenstein, only cognitive impairment was identified as a predictive variable related to the thymic/cognitive [31]. Conversely, Ikeda et al. [28], Kawazoe et al. [27] and Bath et al. [30] identified at least two predictive variables involving the thymic/cognitive dimension, providing a better understanding of the ROF associated with the thymic/cognitive dimension and facilitating preventive measures. In those features, we can find fear of falling, depressive symptoms, self-rated health, impaired consciousness and dementia at admission. Recent studies by Ikeda et al. [28] and Kawazoe et al. [27] achieved Area Under the receiver operating characteristic Curve (AUC) scores of 88% and 85%, respectively, using comprehensive approaches. Ikeda et al. [28] employed a Random Forest-based Boruta algorithm for feature selection, while Kawazoe et al. [27] used a combination of Bidirectional Encoders and Bidirectional Long Short-Term Memory (BiLSTM) networks to process sequential data. These AUC scores indicate strong model performance, reflecting high discriminative ability in classification tasks [25].

Pennone et al. [32] highlighted the difficulty in predicting fall risk among older adults with low levels of daily activity, emphasizing the importance of measuring such activity using standardized indicators. In our predictive model, we included ADL and IADL scores, which are already well-established in the literature as robust predictive factors [33], [34], [35]. A history of falling, which by definition places an older adult at risk of recurrent falls has consistently been identified as a major predictor in recent studies when collected. It is also consistently ranked among the most influential variables in predictive fall models [28], [29], [36], [37]. The cognitive dimension represented here by the MMSE score has also been widely recognized in prior research as an important determinant of fall risk [38], [39], [40]. In addition, Bharadwaz et al. [41] emphasized the influence of depression and sleep disorders on fall risk. Although the pathological GDS score was not among the most influential variables in our final model, it remains relevant when analyzing trajectories. As for sleep disturbances, while not directly measured their impact likely manifests indirectly through reduced performance in activities of daily living further justifying the inclusion of ADL and IADL scores in our predictive approach.

Pathological SLB, combined with the SPPB score, which evaluates gait and balance ability, emerged as one of the strongest determinants in predicting fall risk. Several studies have confirmed that these variables reflecting the physical and organic dimension are essential fall predictors [36], [42], [43], [44]. In the work of Lathouwers et al. [45], it was also shown that maintaining physical, mental, or social activity significantly reduces the probability of falling in older adults, a finding that aligns with our own results.

Indeed, Landers et al. [46] demonstrated that such activities help prevent the onset of fear of falling (FOF) and contribute to maintaining a high level of confidence in one's balance abilities as measured by the Activities-specific Balance Confidence (ABC) scale, both identified as major risk factors. Similarly, Schumann et al. [47] recently highlighted the role of FOF as a predictor of falling.

The only variable present in our model that is notably absent in recent studies is dyslipidemia, a cardiovascular risk factor. This discrepancy may be explained by the methodological specificity of our study, which was based on data collected directly from patients in their homes, allowing for a more integrative assessment of overall health. The inclusion of dyslipidemia in our model underscores the importance of considering cardiovascular risk as a potential contributor to falls, especially when falls occur suddenly and without prior functional warning signs.

While fall history is consistently identified as one of the most influential predictors of future falls, our analysis shows that the model does not rely exclusively on this variable to make its predictions (Fig. 3 and Fig. 4). SHAP value interpretation reveals that the XGBoost model incorporates a wide range of factors, including physical performance, functional autonomy, cognitive status and psychosocial indicators, when estimating fall risk.

In several correctly classified cases, the presence of a prior fall is counterbalanced by protective factors such as high ADL and IADL scores, preserved cognitive function (as indicated by MMSE) and non-pathological balance performance (e.g., SPPB score or SLB). This demonstrates that the model takes into account the complex interplay between risk and protective variables rather than basing its prediction on fall history alone.

Inversely, certain misclassified cases highlight that a history of falls does not always lead to a high-risk prediction. When other variables present a favorable profile, the model may underestimate the actual risk, suggesting that fall history while important is insufficient on its own to ensure predictive accuracy.

Moreover, the model's use of additional variables such as dyslipidemia and cognitive scores reflects a broader more integrative view of fall risk. These results confirm the necessity of a multidimensional approach and support the implementation of interpretable machine learning models that can provide individualized, clinically meaningful insights beyond any single predictor.

This study confirms the relevance of machine learning models, particularly XGBoost for predicting fall risk in older adults with good discriminative performance and calibration. The analysis of SHAP values enabled a transparent and clinically meaningful ranking of predictive factors. Fall history, impairments in physical performance (SPPB, one-leg stance) and functional limitations (ADL, IADL) emerged as the main determinants. Cognitive and psychosocial factors play a secondary yet non-negligible role. These findings highlight the importance of a multidimensional assessment that incorporates interpretable technological tools to guide personalized prevention strategies. The integration of such approaches into geriatric practice could enhance early identification of at-risk patients and contribute to reducing the incidence of falls.

Nonetheless, our work presents several limitations. First, although the XGBoost model demonstrated good performance (AUC of 0.76, Brier score of 0.19, precision of 0.78), its implementation in clinical practice could be hindered by the time required to perform the assessments, even though the number of variables that significantly influence predictions is relatively low. This complexity may limit its use by healthcare professionals in care settings where workload and time constraints are critical factors. A clinical arbitration process aimed at identifying substitutable or priority variables could facilitate the operational integration of the model.

Moreover, the model was built using all variables identified as predictive, without applying a selection procedure based solely on significant differences between fallers and non-fallers. Such a selection approach might optimize the trade-off between predictive performance and ease of use.

From a methodological standpoint, the study did not include a control group. A randomized design comparing a control group (receiving no care) and an intervention group (receiving personalized follow-up) would have allowed for a more detailed analysis of the impact of care on the dynamics of fall risk factors and would have helped to better identify common or distinguishing predictive variables between the two groups.

Finally, the data used were exclusively collected from patients in France. This geographical limitation restricts the generalizability of the findings to other cultural and socio-environmental contexts. Since falls are a multifactorial phenomenon strongly influenced by lifestyle, home environment and care practices, significant variations may exist in other countries. In particular, the socio-environmental dimension deserves to be examined through a multicenter international approach.

Overall, while our model is grounded in a realistic approach aimed at clinical integration, these limitations open avenues for improvement in both methodological robustness and the transferability of results.

V. CONCLUSION

This study contributes to advancing fall prevention by leveraging a 12-year dataset collected in home settings to develop an AI-based predictive model. Our approach integrates the three dimensions of ROF, optimizing model performance while reducing the number of required input features.

By applying explainable AI techniques, we identified the contribution of each feature to fall risk, thereby supporting the development of more targeted and effective intervention strategies. These findings may help enhance the quality of elderly care by informing personalized prevention efforts and guiding future research in geriatric risk assessment.

As with most AI models, ours can be continuously refined with additional data over time. In our case, improving the model also provides an opportunity to collect data from patients' homes while offering them personalized fall prevention advice. During the intervals between practitioner visits, necessary adjustments to home configurations can also be made if needed.

The clinical utility of the final model could be explored in future studies using Decision Curve Analysis (DCA). This method helps identify the clinical range in which the model provides a net benefit, thereby allowing practitioners to determine the optimal threshold for patient management while taking available resources into account.

ACKNOWLEDGMENT

AUTHOR CONTRIBUTIONS

Amadou M. Djiogomaye Ndiaye: Contributed to the study concept and design, drafted the manuscript, conducted statistical analyses, built the models, interpreted the data and approved the final version for publication.

Michel Harel: Contributed to the study concept and design, critically revised the manuscript and approved the final version for publication.

Laurent Billonnet: Contributed to the study concept and design, critically revised the manuscript and approved the final version for publication.

Achille Tchalla: Contributed to the study concept and design, analyzed and interpreted the data, critically revised the manuscript and approved the final version for publication.

SPONSOR'S ROLE

Data grant by *"CARSAT Centre-Ouest",* Limoges, France were collected by UPSAV and URCI. The analysis and manuscript preparation were funded by a grant from chair of excellence in Artificial Intelligence and Healthy Ageing in the Nouvelle-Aquitaine Region (IABV-NA – *Intelligence Artificielle et Bien-Vieillir en Nouvelle-Aquitaine), Limoges, France.*

REFERENCES

[1] A. M. D. Ndiaye, M. Harel, L. Billonnet, and A. Tchalla, "Enhancing Fall Prediction in Older Adults: A Data-Driven Approach to Key Parameter Selection," presented at the eTELEMED 2025, The Seventeenth International Conference on eHealth, Telemedicine, and Social Medicine, May 2025, pp. 37–39. Accessed: Sept. 14, 2025. [Online]. Available: https://www.thinkmind.org/library/eTELEMED/eTELEMED_2025/etelemed_2025_1_80_40047.html

[2] WHO, "Ageing and health." Accessed: Mar. 28, 2023. [Online]. Available: https://www.who.int/fr/news-room/fact-sheets/detail/ageing-and-health

[3] Insee, "Life Expectancy at Various Ages | Insee." Accessed: Mar. 27, 2023. [Online]. Available: https://www.insee.fr/fr/statistiques/2416631#graphique-figure1

[4] X. Thierry, "Accidents and physical assaults among the elderly: less frequent than among the young, but more severe," Population & Sociétés, vol. 468, no. 6, pp. 1–4, 2010, doi: 10.3917/popsoc.468.0001.

[5] H.-C. Tan et al., "Deep learning model for the prediction of all-cause mortality among long term care people in China: a prospective cohort study," Sci Rep, vol. 14, no. 1, p. 14639, June 2024, doi: 10.1038/s41598-024-65601-4.

[6] A. Shaik et al., "A Staged Approach using Machine Learning and Uncertainty Quantification to Predict the Risk of Hip Fracture," ArXiv, p. arXiv:2405.20071v1, May 2024.

[7] US Preventive Services Task Force, "Interventions to Prevent Falls in Community-Dwelling Older Adults: US Preventive Services Task Force Recommendation Statement," JAMA, vol. 319, no. 16, pp. 1696–1704, Apr. 2018, doi: 10.1001/jama.2018.3097.

[8] Kellogg, "The prevention of falls in later life. A report of the Kellogg International Work Group on the Prevention of Falls by the Elderly," Dan Med Bull, vol. 34 Suppl 4, pp. 1–24, Apr. 1987.

[9] S. L. Wolf, H. X. Barnhart, N. G. Kutner, E. McNeely, C. Coogler, and T. Xu, "Reducing frailty and falls in older persons: an investigation of Tai Chi and computerized balance training. Atlanta FICSIT Group. Frailty and Injuries: Cooperative Studies of Intervention Techniques," Journal of the American Geriatrics Society, vol. 44, no. 5, May 1996, doi: 10.1111/j.1532-5415.1996.tb01432.x.

[10] A. Pilotto et al., "Three Decades of Comprehensive Geriatric Assessment: Evidence Coming From Different Healthcare Settings and Specific Clinical Conditions," Journal of the American Medical Directors Association, vol. 18, no. 2, p. 192.e1-192.e11, Feb. 2017, doi: 10.1016/j.jamda.2016.11.004.

[11] M. E. Tinetti and M. Speechley, "Prevention of Falls among the Elderly," http://dx.doi.org/10.1056/NEJM198904203201606. Accessed: Apr. 24, 2023. [Online]. Available: https://www.nejm.org/doi/pdf/10.1056/NEJM198904203201606

[12] A. L. Benton, "Development of a multilingual aphasia battery: Progress and problems," Journal of the Neurological Sciences, vol. 9, no. 1, pp. 39–48, July 1969, doi: 10.1016/0022-510X(69)90057-4.

[13] T. H. Trojian and D. B. McKeag, "Single leg balance test to identify risk of ankle sprains," July 2006, doi: 10.1136/bjsm.2005.024356.

[14] M. Freedman, L. Leech, E. Kaplan, G. Winocur, K. Shulman, and D. Delis, "M. Freedman, L. Leech, E. Kaplan, G. Winocur, K. Shulman and D. Delis. Clock drawing: a neuropsychological analysis. New York: Oxford Press, 1994. | Request PDF," ResearchGate, 1994, doi: 10.1017/S0714980800013398.

[15] M. P. Lawton and E. M. Brody, "Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living1," The Gerontologist, vol. 9, no. 3_Part_1, pp. 179–186, Oct. 1969, doi: 10.1093/geront/9.3_Part_1.179.

[16] J. I. Escobar, A. Burnam, M. Karno, A. Forsythe, J. Landsverk, and J. M. Golding, "Use of the Mini-Mental State Examination (MMSE) in a community population of mixed ethnicity. Cultural and linguistic artifacts," J Nerv Ment Dis, vol. 174, no. 10, pp. 607–614, Oct. 1986, doi: 10.1097/00005053-198610000-00005.

[17] B. Vellas et al., "The Mini Nutritional Assessment (MNA) and its use in grading the nutritional state of elderly patients," Nutrition, vol. 15, no. 2, pp. 116–122, Feb. 1999, doi: 10.1016/s0899-9007(98)00171-3.

[18] J. M. Guralnik et al., "A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission," J Gerontol, vol. 49, no. 2, pp. M85-94, Mar. 1994, doi: 10.1093/geronj/49.2.m85.

[19] J. A. Yesavage et al., "Development and validation of a geriatric depression screening scale: A preliminary report," Journal of Psychiatric Research, vol. 17, no. 1, pp. 37–49, Jan. 1982, doi: 10.1016/0022-3956(82)90033-4.

[20] W. S. Noble, "What is a support vector machine?," Nat Biotechnol, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.

[21] L. Breiman, Classification and Regression Trees. New York: Routledge, 2017. doi: 10.1201/9781315139470.

[22] J. Brownlee, XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn. Machine Learning Mastery, 2016.

[23] T. Schlosser, M. Friedrich, T. Meyer, and D. Kowerko, "A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision," ResearchGate. Accessed: May 28, 2025. [Online]. Available: https://www.researchgate.net/publication/374558675_A_Consolidated_Overview_of_Evaluation_and_Performance_Metrics_for_Machine_Learning_and_Computer_Vision

[24] L. Hoessly, "On misconceptions about the Brier score in binary prediction models," Apr. 23, 2025, arXiv: arXiv:2504.04906. doi: 10.48550/arXiv.2504.04906.

[25] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recognition, vol. 30, no. 7, pp. 1145–1159, July 1997, doi: 10.1016/S0031-3203(96)00142-2.

[26] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. Accessed: May 14, 2025. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

[27] Y. Kawazoe, K. Shimamoto, D. Shibata, E. Shinohara, H. Kawaguchi, and T. Yamamoto, "Impact of a Clinical Text-Based Fall Prediction Model on Preventing Extended Hospital Stays for Elderly Inpatients: Model Development and Performance Evaluation," JMIR Med Inform, vol. 10, no. 7, p. e37913, July 2022, doi: 10.2196/37913.

[28] T. Ikeda et al., "An Interpretable Machine Learning Approach to Predict Fall Risk Among Community-Dwelling Older Adults: a Three-Year Longitudinal Study," J GEN INTERN MED, vol. 37, no. 11, pp. 2727–2735, Aug. 2022, doi: 10.1007/s11606-022-07394-8.

[29] A. Cella et al., "Development and validation of a robotic multifactorial fall-risk predictive model: A one-year prospective study in community-dwelling older adults," PLOS ONE, vol. 15, no. 6, p. e0234904, juin 2020, doi: 10.1371/journal.pone.0234904.

[30] P. A. Bath, N. Pendleton, K. Morgan, J. E. Clague, M. A. Horan, and S. B. Lucas, "New approach to risk determination: development of risk profile fornew falls among community-dwelling older people by use of a GeneticAlgorithm Neural Network (GANN)," The Journals of Gerontology: Series A, vol. 55, no. 1, pp. M17–M21, Jan. 2000, doi: 10.1093/gerona/55.1.M17.

[31] L. Z. Rubenstein, "Falls in older people: epidemiology, risk factors and strategies for prevention," Age and Ageing, vol. 35, no. suppl_2, pp. ii37–ii41, Sept. 2006, doi: 10.1093/ageing/afl084.

[32] J. Pennone, N. F. Aguero, D. M. Martini, L. Mochizuki, and A. A. do Passo Suaide, "Fall prediction in a quiet standing balance test via machine learning: Is it possible?," PLoS One, vol. 19, no. 4, p. e0296355, 2024, doi: 10.1371/journal.pone.0296355.

[33] W.-M. Chu 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., vol. 9, 2022, doi: 10.3389/fmed.2022.937216.

[34] C.-W. Kang, Z.-K. Yan, J.-L. Tian, X.-B. Pu, and L.-X. Wu, "Constructing a fall risk prediction model for hospitalized patients using machine learning," BMC Public Health, vol. 25, no. 1, p. 242, Jan. 2025, doi: 10.1186/s12889-025-21284-8.

[35] A. K. Mishra et al., "Explainable Fall Risk Prediction in Older Adults Using Gait and Geriatric Assessments," Front. Digit. Health, vol. 4, 2022, doi: 10.3389/fdgth.2022.869812.

[36] S. Chen et al., "Comparing interpretable machine learning models for fall risk in middle-aged and older adults with and without pain," Sci Rep, vol. 15, no. 1, p. 17032, May 2025, doi: 10.1038/s41598-025-01651-6.

[37] L. Lin, X. Liu, C. Cai, Y. Zheng, D. Li, and G. Hu, "Urban-rural disparities in fall risk among older Chinese adults: insights from machine learning-based predictive models," Front Public Health, vol. 13, p. 1597853, 2025, doi: 10.3389/fpubh.2025.1597853.

[38] O. Beauchet 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, vol. 22, no. 1, pp. 131–137, 2018, doi: 10.1007/s12603-017-0950-z.

[39] T. Ikeda et al., "An Interpretable Machine Learning Approach to Predict Fall Risk Among Community-Dwelling Older Adults: a Three-Year Longitudinal Study," J Gen Intern Med, vol. 37, no. 11, pp. 2727–2735, Aug. 2022, doi: 10.1007/s11606-022-07394-8.

[40] A. Kabeshova et al., "Falling in the elderly: Do statistical models matter for performance criteria of fall prediction? Results from two large population-based studies," Eur. J. Intern. Med., vol. 27, pp. 48–56, 2016, doi: 10.1016/j.ejim.2015.11.019.

[41] M. P. Bharadwaz, J. Kalita, A. Mitro, and A. Aditi, "Utilizing machine learning to identify fall predictors in India's aging population: findings from the LASI," BMC Geriatr, vol. 25, no. 1, p. 181, Mar. 2025, doi: 10.1186/s12877-025-05813-z.

[42] G. Cuaya-Simbro, A.-I. Perez-Sanpablo, A. Munõz-Meléndez, I. Q. Uriostegui, E.-F. Morales-Manzanares, and L. Nuñez-Carrera, "Comparison of Machine Learning Models to Predict Risk of Falling in Osteoporosis Elderly," Found. Comput. Decis. Sci., vol. 45, no. 2, pp. 66–77, 2020, doi: 10.2478/fcds-2020-0005.

[43] T. Deschamps, C. G. Le Goff, G. Berrut, C. Cornu, and J.-B. Mignardot, "A decision model to predict the risk of the first fall onset," Experimental Gerontology, vol. 81, pp. 51–55, Aug. 2016, doi: 10.1016/j.exger.2016.04.016.

[44] A. K. Mishra et al., "Explainable Fall Risk Prediction in Older Adults Using Gait and Geriatric Assessments," Front Digit Health, vol. 4, p. 869812, May 2022, doi: 10.3389/fdgth.2022.869812.

[45] E. Lathouwers et al., "Characterizing fall risk factors in Belgian older adults through machine learning: a data-driven approach," BMC Public Health, vol. 22, no. 1, p. 2210, Nov. 2022, doi: 10.1186/s12889-022-14694-5.

[46] M. R. Landers, S. Oscar, J. Sasaoka, and K. Vaughn, "Balance confidence and fear of falling avoidance behavior are most predictive of falling in older adults: prospective analysis," Physical therapy, vol. 96, no. 4, pp. 433–442, 2016, Accessed: June 17, 2024. [Online]. Available: https://academic.oup.com/ptj/article-abstract/96/4/433/2686463

[47] P. Schumann et al., "Using machine learning algorithms to detect fear of falling in people with multiple sclerosis in standardized gait analysis," Mult Scler Relat Disord, vol. 88, p. 105721, Aug. 2024, doi: 10.1016/j.msard.2024.105721.