

# An Interpretable Machine Learning Approach to Predict Fall Risk Among Community-Dwelling Older Adults: a Three-Year Longitudinal Study



Takaaki Ikeda, RPT, PhD<sup>1,2</sup> , Upul Cooray, MSc<sup>2</sup> , Masanori Hariyama, PhD<sup>3</sup>, Jun Aida, PhD<sup>4,5</sup> , Katsunori Kondo, PhD<sup>6,7</sup>, Masayasu Murakami, PhD<sup>1</sup> , and Ken Osaka, PhD<sup>2</sup>

<sup>1</sup>Department of Health Policy Science, Graduate School of Medical Science, Yamagata University, Yamagata, Japan; <sup>2</sup>Department of International and Community Oral Health, Graduate School of Dentistry, Tohoku University, Sendai, Miyagi, Japan; <sup>3</sup>Intelligent Integrated Systems Laboratory, Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan; <sup>4</sup>Department of Oral Health Promotion, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo, Japan; <sup>5</sup>Division for Regional Community Development, Liaison Center for Innovative Dentistry, Graduate School of Dentistry, Tohoku University, Sendai, Miyagi, Japan; <sup>6</sup>Department of Social Preventive Medical Sciences, Center for Preventive Medical Sciences, Chiba University, Chiba, Japan; <sup>7</sup>Department of Gerontological Evaluation, Center for Gerontology and Social Science, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan.

**BACKGROUND:** Adverse health effects resulting from falls are a major public health concern. Although studies have identified risk factors for falls, none have examined long-term prediction of fall risk. Furthermore, recent evidence suggests that there are additional risk factors, such as psychosocial factors.

**OBJECTIVE:** In this 3-year longitudinal study, we evaluated a predictive model for risk of fall among community-dwelling older adults using machine learning methods.

**DESIGN:** A 3-year follow-up prospective longitudinal study (from 2010 to 2013).

**SETTING:** Twenty-four municipalities in nine of the 47 prefectures (provinces) of Japan.

**PARTICIPANTS:** Community-dwelling individuals aged  $\geq 65$  years who were functionally independent at baseline ( $n = 61,883$ ).

**METHODS:** The baseline survey was conducted from August 2010 to January 2012, and the follow-up survey was conducted from October to December 2013. Both surveys were conducted involving self-reported questionnaires. The measured outcome at the follow-up survey was self-reported multiple falls during the previous year. The 142 variables included in the baseline survey were regarded as candidate predictors. The random-forest-based Boruta algorithm was used to select predictors, and the eXtreme Gradient Boosting algorithm with 10 repetitions of nested  $k$ -fold cross-validation was used for modeling and model evaluation. Furthermore, we used shapley additive explanations to gain insight into the behavior of the prediction model.

**KEY RESULTS:** Fourteen out of 142 candidate features were selected as predictors. Among these predictors, experience of falling as of the baseline survey was the most important feature, followed by self-rated health and age. Moreover, sense of coherence was newly identified as a risk factor for falls.

**CONCLUSIONS:** This study suggests that machine learning tools can be adapted to explore new associative factors, make accurate predictions, and provide actionable insights for fall prevention strategies.

**KEY WORDS:** Boruta; eXtreme Gradient Boosting; fall prediction; psychosocial factors; random forest.

J Gen Intern Med 37(11):2727–35  
DOI: 10.1007/s11606-022-07394-8

© The Author(s) under exclusive licence to Society of General Internal Medicine 2022

## INTRODUCTION

The cascade of adverse health effects resulting from falls is an important public health concern worldwide.<sup>1,2</sup> Although the age-standardized incidence rate of falls was less in 2017 than in 1990,<sup>1</sup> falls still account for a large proportion of injuries in older adults.<sup>1,2</sup> A recent study reported that approximately 30% of older adults experienced falls annually and that fallers were more likely to suffer from disability due to fall-related injuries.<sup>3</sup> Thus, fall prevention is important to prevent the deterioration of personal health, including physical and mental health, and to reduce social burdens, such as healthcare costs.<sup>4</sup>

Multiple factors are considered to contribute to fall risk. Systematic reviews have reported that poor physical function,<sup>5,6</sup> cognitive decline,<sup>7</sup> malnutrition,<sup>8</sup> obesity,<sup>8,9</sup> physical inactivity,<sup>10</sup> polypharmacy,<sup>11</sup> and urinary incontinence<sup>12</sup> are risk factors for falls among older adults. In addition to these risk factors, psychosocial factors<sup>13–17</sup> have gained increasing attention as important factors. These include the fear of falling,<sup>13</sup> socioeconomic status,<sup>14–16</sup> social participation,<sup>16</sup> and fall efficacy.<sup>17</sup> These studies suggest that there may still be unknown risk factors, related in particular to psychosocial aspects. In addition, because existing studies do not directly compare the various risk factors, the risk factors that are most likely to predict falls in older adults remain unclear.

---

Takaaki Ikeda and Upul Cooray contributed equally.

Received May 3, 2021

Accepted January 3, 2022

Published online February 2, 2022

In recent years, studies using machine learning approaches have gained increasing attention.<sup>18–22</sup> However, previous machine learning studies of fall risk have failed to obtain an accurate predictive model.<sup>21,22</sup> This situation may be explained by the following reasons: (1) Frailty was more prevalent in each previous study's target population than in the average population of the studied age range,<sup>21,22</sup> and the risk of falling is relatively elevated in frail, older people.<sup>23</sup> Thus, the predictors used in their study may have had poor predictive ability for future fall risk because the characteristics of the fallers and nonfallers in those studies were similar. (2) Previous studies lacked information on the variables related to psychosocial aspects owing to the nature of the medical claim data. In general, medical claim data mainly involve physical data, including diagnosis, medication use, and physical examination data; therefore, they lack psychosocial information. (3) The machine learning algorithm used in previous studies (random forest) were less powerful, leading to lower performance metrics. In general, eXtreme Gradient Boosting (XGBoost) algorithm,<sup>24</sup> a new algorithm, performs better than random forest algorithm<sup>25</sup> owing to the former's serial learning algorithm that builds trees one at a time, where a new tree helps to correct the errors made by the previously trained tree. Meanwhile, the random forest algorithm builds each tree independently.<sup>25</sup>

Machine learning approaches as a whole are advancing rapidly. In addition, new frameworks are becoming available to make complex machine learning models more interpretable, enabling the extraction of actionable insights from models.<sup>26,27</sup> Therefore, the exploration of new predictors and their behavior in predicting falls with an interpretable machine learning approach can help in developing preventive strategies for community-dwelling older adults. Recently, psychosocial factors have been considered especially important for older adults because of these factors' bidirectional relationships.<sup>13–17,28</sup> For example, a longitudinal study targeting the general population reported that depressive symptoms and weak handgrip strength were mutually associated with one another.<sup>28</sup>

In this study, we take advantage of the Japan Gerontological Evaluation Study (JAGES), which conducted a large-scale ongoing panel survey targeted at older adults across Japan using a multidimensional questionnaire. The JAGES questionnaire involves multidimensional variables covered in the surveys, including health, psychological, and functional factors.<sup>29,30</sup> We aim to construct an interpretable predictive machine learning model for falls among community-dwelling older adults using machine learning methods.

## METHODS

### Study Population

We constructed panel data from the JAGES, which is an ongoing nationwide study targeting community-dwelling older adults in Japan. The baseline survey was conducted from

August 2010 to January 2012. Self-reported questionnaires were mailed to 141,407 individuals aged  $\geq 65$  years who were not certified in the long-term care insurance system across 24 municipalities in nine of the 47 prefectures (provinces) of Japan. A follow-up survey was conducted from October to December 2013, and self-administered questionnaires for the follow-up survey were mailed to the same respondents (Figure 1).

### Variables

The experience of falls during the past year was used as the participant's outcome of interest. Self-reported falls were evaluated by asking the participants, "Have you had any falls over the past year?" with possible responses of "multiple times," "once," or "none." Participants who provided the latter two responses were combined as previous studies have reported that annual fallers exhibited health characteristics similar to those of nonfallers.<sup>31,32</sup>

First, variables with more than 30% missing information were excluded from the analysis.<sup>33</sup> We used all the remaining 142 variables included in the baseline survey of the JAGES as candidate predictors (Supplementary Table 1). Some baseline variables were aggregated to calculate the scores of the following corresponding scales: Japanese Geriatric Depression Scale,<sup>34</sup> Tokyo Metropolitan Institute of Gerontology Index of Competence,<sup>35</sup> and sense of coherence (SOC) scale (Supplementary Table 2).<sup>36</sup> We also calculated the study period and included as a candidate feature.

In the present study, the Japanese Geriatric Depression Scale was a 15-item scale whose total score was calculated by adding negative responses ranging from 0 to 15, with higher scores indicating a higher probability of depression. We categorized respondents into the following three groups: no depression (0–4 points), mild depression (5–9 points), and severe depression (10–15 points).<sup>34</sup>

The Tokyo Metropolitan Institute of Gerontology Index of Competence is a 13-item scale (in yes/no format) to assess the abilities regarding physical functions, effectance, and social roles.<sup>35</sup> In this scale, a high score indicates high ability.<sup>35</sup> Each of the three domains was aggregated separately and used as a candidate feature.

In this study, SOC was measured using six questions (two questions from each of the three subdomains employed in the SOC Scale).<sup>36</sup> SOC, which is defined as the ability to cope with stressful life experiences, is considered to reflect personal health behavior.<sup>37</sup> According to a previous study,<sup>38</sup> the responses were summed to create a score that ranged from 6 to 30, with a high score indicating a high level of SOC. We categorized the respondents into the following three groups, in keeping with a previous study: low SOC (6–20 points), middle SOC (21–23 points), and high SOC (24–30 points).<sup>38</sup>

### Analytic Strategy

The random forest imputation algorithm was used in this study to handle missing variables.<sup>33</sup> The variables that were assessed

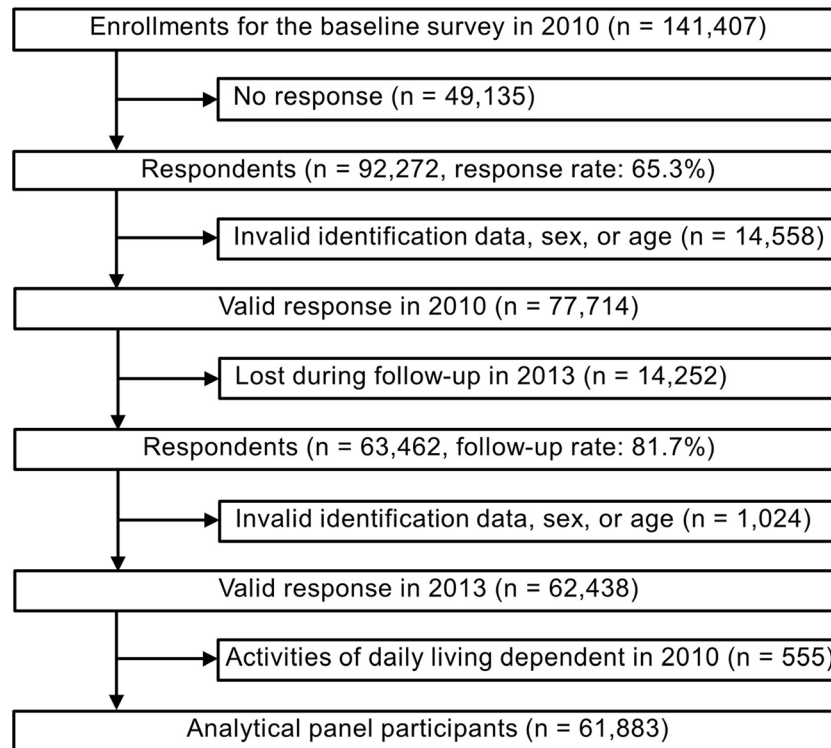


Figure 1 Participant selection for analysis, Japan, 2010–2013.

in the follow-up survey were used as the explanatory variables in addition to the 142 candidate features (measured in the baseline survey) in our imputation procedure, and the follow-up variables other than our outcome (i.e., experience of falls in the past year) were excluded after the imputation procedure.

Our machine learning procedure comprised the three steps: feature selection, modeling, and model evaluation and SHapley Additive exPlanations (SHAP) value calculation. In a high-dimensional dataset such as ours, selection of features is an important procedure in machine learning to improve interpretability, avoid overfitting, and prevent performance degradation.<sup>39</sup> The random-forest-based Boruta algorithm was used for feature selection.<sup>39</sup> This algorithm is reported to be one of the most robust feature selection algorithms and recommended for the analysis of high-dimensional dataset.<sup>40</sup> For modeling and model evaluation, we used a nested  $k$ -fold cross-validation procedure to prevent overfitting and overly optimistic estimates of model performance.<sup>41,42</sup> First, the dataset was randomly split into 10 mutually exclusive folds (outer split). Nine of the 10 folds were used as *training data* to train the model, while the remaining fold was used as *test data* for model evaluation. The training data from the outer split were further divided into 5 folds (inner split). Four of these 5 folds were used for hyperparameter optimization (training set) and validated with the remaining fold (validation set). This process was iterated until each fold in the outer split was evaluated as a test (10 iterations). Then, the entire nested  $k$ -

fold cross-validation procedure was repeated 10 times, evaluating 100 independent models (Figure 2).

We used a random search strategy to identify the optimal hyperparameters for the training model<sup>43</sup> and applied the XGBoost algorithm, which is based on a decision tree framework, for the learning procedure.<sup>24</sup> Mean performance metrics were calculated from the 100 evaluated models. Model performance metrics included accuracy score, F1 score, and area under the receiver operating characteristic curve (AUC). The formulas for each performance metric are summarized in Supplementary Figure 1.

The combination of an oversampling method (synthetic minority oversampling technique) and under-sampling method (edited nearest neighbor)<sup>44</sup> was applied to handle the imbalanced class distribution of the data.<sup>45,46</sup> The sampling method we used in this study forms new samples using  $k$ -nearest neighbors and then cleans the oversampled data.<sup>44</sup>

To compare the magnitude of the contribution of each predictor, we computed the SHAP value, which is a novel framework based on the game theory.<sup>27</sup> The SHAP value quantifies the contribution of each feature to the prediction results. This framework allowed us to verify whether each feature contributed positively or negatively to the probability of falling. We calculated SHAP values for the XGBoost model with the highest predictive capacity (i.e., highest AUC).

For comparison, we calculated the prediction performance scores with imbalanced data and with random under-sampling with a 1:1 ratio. Furthermore, we used random forest models<sup>25</sup>

Table 1 Baseline characteristics according to experience of falls in follow-up survey, Japan, 2010–2013

Characteristics	Total	Nonfallers/annual fallers		Multiple fallers		Statistics <i>P</i> -value
		<i>n</i>	%	<i>n</i>	%	
Age (years)						
65–69	20,714	20,036	34.2	678	20.2	<0.001
70–74	18,945	18,071	30.9	874	26.0	
75–79	13,554	12,622	21.6	932	27.8	
80–84	6,570	5,940	10.2	630	18.8	
≥85	2,100	1,855	3.2	245	7.3	
Sex (Female)	33,242	31,459	53.8	1,783	53.1	0.45
Education level (years)						
<6	1,012	889	1.5	123	3.7	<0.001
6–9	28,453	26,599	45.5	1,854	55.2	
10–12	21,463	20,521	35.1	942	28.0	
≥13	10,955	10,515	18.0	440	13.1	
Equalized household income						
First (lowest)	13,628	12,651	21.6	977	29.1	<0.001
Second	12,344	11,589	19.8	755	22.5	
Third	16,434	15,688	26.8	746	22.2	
Fourth	9,384	8,954	15.3	430	12.8	
Fifth (highest)	10,093	9,642	16.5	451	13.4	
Experience of falling						
Multiple times	3,395	2,233	3.8	1,162	34.6	<0.001
Once	13,297	12,208	20.9	1,089	32.4	
None	45,191	44,083	75.3	1,108	33.0	
Self-rated health						
Poor	1,027	859	1.5	168	5.0	<0.001
Fair	8,663	7,695	13.2	968	28.8	
Good	44,499	42,451	72.5	2,048	61.0	
Excellent	7,694	7,519	12.9	175	5.2	
Fear of falling (yes)	25,899	23,733	40.6	2,166	64.5	<0.001
Ability to stand up from chairs (no)	7,889	6,945	11.9	944	28.1	<0.001
Depressive symptoms						
None	47,911	45,906	78.4	2,005	59.7	<0.001
Mild	10,683	9,788	16.7	895	26.6	
Severe	3,289	2,830	4.8	459	13.7	
Choking (yes)	8,726	7,808	13.3	918	27.3	<0.001
Dry mouth (yes)	11,168	10,105	17.3	1,063	31.7	<0.001
Having arthrosis (yes)	6,470	5,784	9.9	686	20.4	<0.001
Difficulties in eating tough foods (yes)	14,697	13,426	22.9	1,271	37.8	<0.001
Ability to climb stairs (no)	21,656	19,887	34.0	1,769	52.7	<0.001
Sense of coherence						
Low	21,697	19,989	34.2	1,708	50.9	<0.001
Moderate	25,820	24,671	42.2	1,149	34.2	
High	14,366	13,864	23.7	502	14.9	
Incontinence (yes)	4,061	3,592	6.1	469	14.0	<0.001
Number of remaining teeth						
≥20	23,344	22,534	38.5	810	24.1	<0.001
10–19	15,860	15,024	25.7	836	24.9	
1–9	15,436	14,360	24.5	1,076	32.0	
Edentulous	7,243	6,606	11.3	637	19.0	

<sup>a</sup>Chi-square test was performed

as a baseline for performance comparison with XGBoost models. We also implemented a conventional logistic

regression model using the selected features via the random-forest-based Boruta algorithm. All analyses were performed

Table 2 Prediction performance of XGBoost and random forest algorithms with and without resampling

Model	Resampling	Accuracy		F1 score		AUC	
		Mean	SD	Mean	SD	Mean	SD
XGBoost	Adopted combination of oversampling and under-sampling	0.88	0.02	0.89	0.02	0.88	0.02
	Adopted under-sampling	0.77	0.01	0.72	0.01	0.72	0.01
	Not adopted (imbalanced)	0.83	0.03	0.25	0.01	0.69	0.03
Random forest	Adopted combination of oversampling and under-sampling	0.86	0.01	0.87	0.01	0.85	0.01
	Adopted under-sampling	0.73	0.01	0.73	0.01	0.73	0.01
	Not adopted (imbalanced)	0.92	0.00	0.18	0.02	0.56	0.01

AUC area under the receiver operating characteristic curve, SD standard deviation, XGBoost eXtreme Gradient Boosting

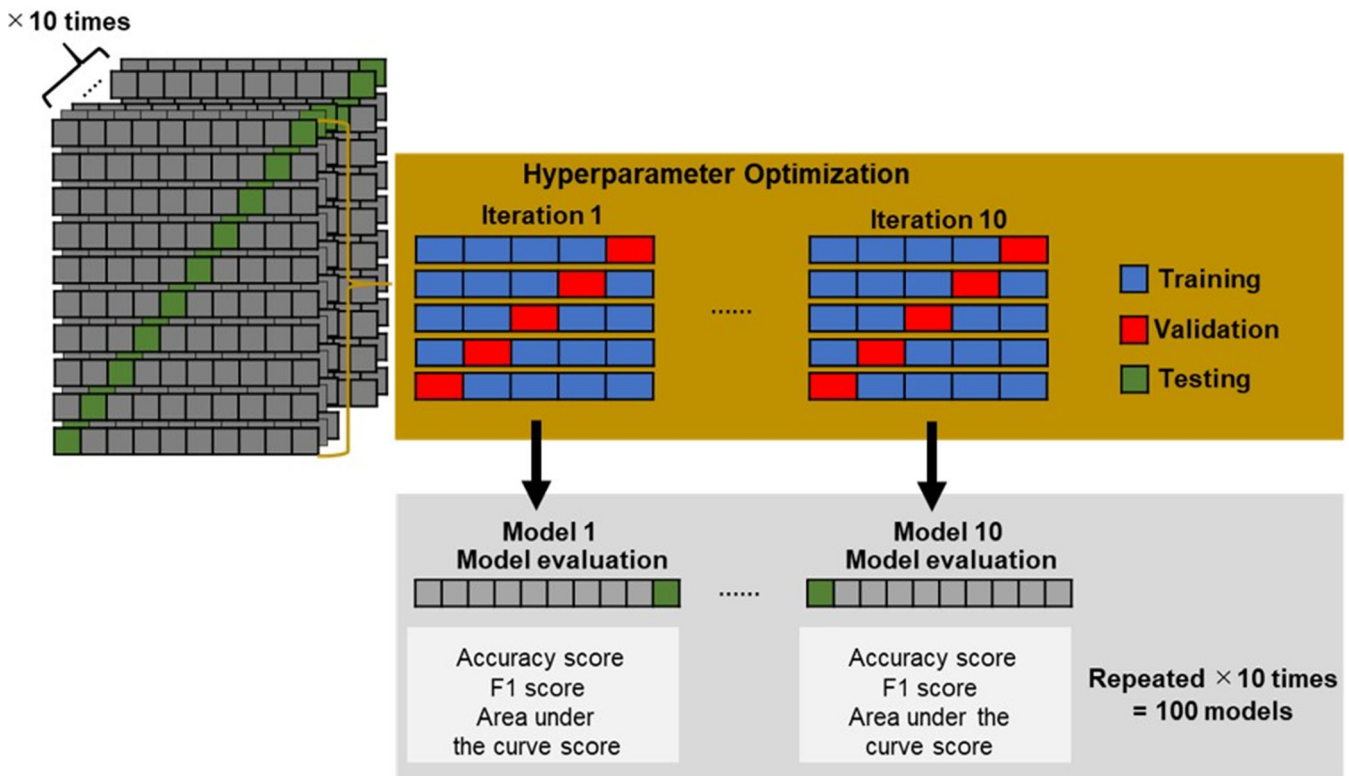


Figure 2 Modeling step of proposed machine learning approach (10 repetitions of nested *k*-fold cross-validation).

using Python 3.8.3. The study protocol was reviewed and approved by ethics committees at Tohoku University.

## RESULTS

### Characteristics of the Analyzed Participants

At the baseline, 92,272 individuals responded to the questionnaire (65.3% response rate). We excluded the participants with invalid baseline information ( $n = 14,558$ ); thus, 77,714 participants remained. In total, 63,462 participants responded to the follow-up survey (81.7% follow-up rate). We then excluded participants who had invalid follow-up information ( $n = 1,024$ ) or were functionally dependent at baseline ( $n = 555$ ). Consequently, 61,883 participants were analyzed in our main analysis (Figure 1). The mean ages of the nonfaller/annual faller group and the multiple faller group were 72.8 (SD = 5.5) and 75.4 (SD = 6.1), respectively. The baseline demographic characteristics of nonfallers/annual fallers and multiple fallers in the follow-up survey are summarized in Table 1.

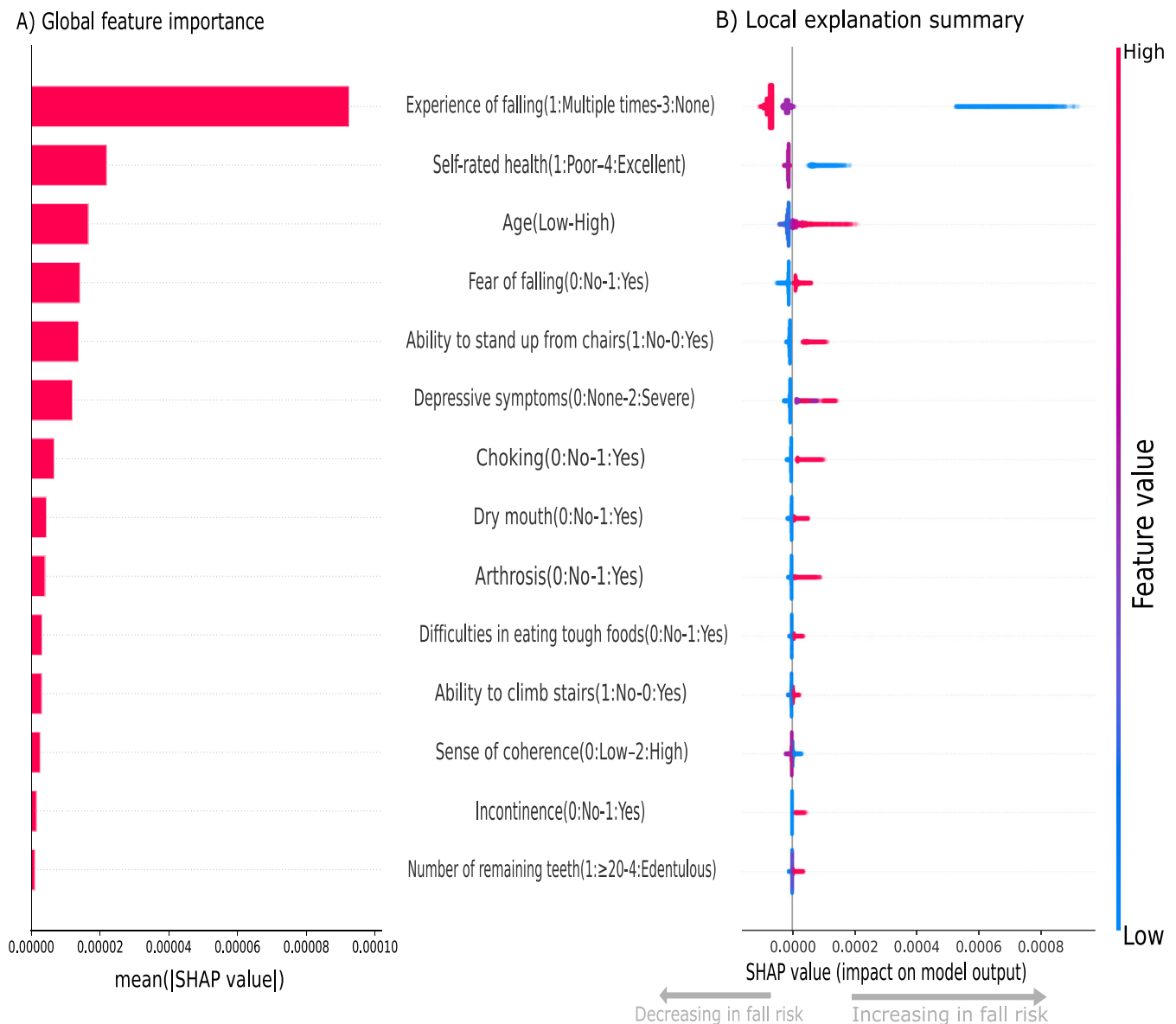
### Selected Features

Fourteen features measured at the time of the baseline survey were selected from the 142 candidates based on the random-forest-based Boruta algorithm (Figure 3): the experience of falling (multiple times, once, or none), self-rated health (poor, fair, good, or excellent), age (continuous), fear of falling (no or

yes), ability to stand up from chairs without using one’s hands (no or yes), depressive symptoms (mild, moderate, or severe), choking experience (no or yes), dry mouth (no or yes), arthritis (no or yes), difficulties in eating hard foods (no or yes), ability to climb stairs without a handrail (no or yes), SOC (low, middle, or high), incontinence (no or yes), and number of remaining teeth ( $\geq 20$ , 10–19, 1–9, or edentulous).

**Prediction of Fall Risk.** The mean prediction performance of our primary model in terms of accuracy, F1 score, and AUC score was 0.88 (SD = 0.02), 0.89 (SD = 0.02), and 0.88 (SD = 0.02), respectively. The scores of each of the 100 models are presented in Supplementary Table 3. Among the predictive scores obtained from the application of an under-sampling method to the imbalanced sample, those obtained from our primary model were the best (Table 2). Moreover, XGBoost obtained better model evaluation scores than the random forest algorithm (Table 2). The implementation of the conventional logistic regression model using the selected features produced results, which are presented in Supplementary Table 4.

Figure 3 presents the calculated SHAP values of our primary model. Figure 3(A) presents the global feature importance, demonstrating that the experience of falling as of the baseline survey was the most important feature, followed by self-rated health and age. Figure 3(B) presents the behavior of each local prediction, where each dot represents an individual prediction. Each dot reveals the direction of effects for different levels of each predictor; for example, the lower values of experience of



**Figure 3 Feature importance and SHapley Additive exPlanations (SHAP) values for each selected feature. (A) Global feature importance. (B) Local explanation summary. The behavior of each local prediction and each dot represents an individual prediction. Each dot represents the direction of effects (positive = red, negative = blue) at different levels of each predictor. When multiple dots are in the same  $x$  position, they accumulate to represent density.**

falling (i.e., multiple fallers; blue dot) were associated with a higher risk of falling than higher values of the experience of falling (i.e., none; red dot).

## DISCUSSION

This study is the first to successfully determine long-term predictors of fall risk among community-dwelling older adults using a machine learning approach. In the present study, 14 features were selected as predictors of falls in the older adults, and SOC was selected as an important new potential predictor.

Many previous systematic reviews have identified risk factors for falling, such as physical functions.<sup>5–12</sup> In addition to these findings, old age,<sup>47</sup> fear of falling,<sup>13</sup> depression,<sup>5</sup> urinary

incontinence,<sup>12</sup> and low self-rated health<sup>48</sup> were reported as risk factors for falling. In keeping with these studies, age, fear of falling, depressive symptoms, self-rated health, and physical function were selected as important features in this study.

In our study, predictors considered to reflect psychosocial aspects were observed as the important feature for fall prediction (Figure 3(A)). For example, we found that depressive symptoms were important predictors of fall risk. The possible mechanism underlying the association between depressive symptoms and physical functions may be explained by the physical inactivity and unintentional weight loss due to depressive symptoms: a previous study reported that individuals with depressive symptoms were more likely to be physically inactive.<sup>49</sup> Moreover, other studies reported that depression is associated with anorexia and weight loss in older people.<sup>50,51</sup>

Furthermore, both physical inactivity and unintentional weight loss are associated with poor physical function.<sup>52,53</sup> Additionally, SOC, which is defined as the ability to cope with stressful life experiences,<sup>37</sup> was also selected as one of the important features among the 142 candidates, which has not been reported in previous studies. In general, individuals with high SOC are able to self-evaluate their social roles by accepting, approving, coping with, and even changing them as required. These individuals are able to engage in specific, sensible, and rational behaviors, and this ability may contribute to healthier behavior.<sup>36</sup> In addition, a longitudinal study reported that lower SOC was a predictor for the onset of depression.<sup>54</sup> Indeed, our conventional regression model showed that the association of SOC with fall risk was somewhat attenuated by depressive symptoms (Supplementary Table 4). This suggests that a low level of SOC predicts the 3-year fall risk by increasing depressive symptoms. However, future studies are warranted to examine the possible mechanism underlying the associations between psychosocial factors and fall risk.

Although further studies investigating those psychosocial factors' contributions to fall risk are warranted, we considered the assessment of psychosocial factors in clinical practice is important for a number of reasons. The relationship between psychosocial factors and physical functions has recently been reported to be bidirectional. For example, a longitudinal study targeting the general population in China reported that depressive symptoms were associated with weak handgrip strength.<sup>28</sup> Moreover, a longitudinal study targeting older women in the USA reported that depressive symptoms were associated with new-onset physical frailty.<sup>55</sup> Therefore, psychosocial factors could be theoretically associated with falling because physical function is one of the risk factors for falling.<sup>5,6</sup>

An observational study that analyzed medical claim data of older patients receiving home healthcare services obtained a moderately accurate predictive model for fall risk: the AUC value of their best model was 0.67.<sup>21</sup> Another previous study that analyzed community-dwelling older adults who were at elevated risk for disability also reported a moderately accurate prediction model: the AUC value of their best model was 0.66.<sup>22</sup> The poorer performance metrics in those previous studies compared with ours might be explained by the following: Frailty was more prevalent in each previous study's target population than in the average population of the studied age range,<sup>21,22</sup> and the risk of falling is relatively elevated in frail, older people.<sup>23</sup> Thus, we considered that within the frail population, previous experience of falling may not accurately predict future falls. Meanwhile, our results showed that previous experience of falling was the strongest predictor of falling. This finding suggests that this is the most important information for assessing fall risk in functionally independent, community-dwelling, older people. The set of candidate features for predicting falls in previous studies was biased toward

somatic or functional information and did not fully capture psychosocial factors, such as mental health or socioeconomic status, which are considered more important in older adults.<sup>13-17</sup> By contrast, among the 14 features of the 142 candidate features in our study, three features with psychosocial factors were selected as important predictors for risk. These particular features were not captured in previous studies that conducted predictive models using machine learning approaches.<sup>21,22</sup> The machine learning algorithm used in previous studies (random forest) was less powerful, leading to lower performance metrics. In general, the XGBoost method used in this study outperforms the random forest method used in previous studies.<sup>21,22</sup> Thus, we consider that the XGBoost method used in our predictive model for fall risk in community-dwelling older adults is more powerful than the random forest method.

This study has several limitations. First, we cannot infer causality from the results, as the use of machine learning does not protect against bias due to unmeasured confounders. Nevertheless, a study has argued that highly accurate predictive models perform better than traditional parametric causal models.<sup>56</sup> Second, most of the features were time-variant variables, and we used only those variables that were assessed in the baseline survey. Third, we could not identify the participants who experienced falling during the intermediate period. Therefore, our estimates of fall risk may underestimate the actual risk. Fourth, the JAGES respondents are not nationally representative. Thus, the difference in the demographic characteristics between the target population and our study population should be considered when applying the current results to other populations.

In conclusion, our machine learning approaches for 3-year fall prediction yielded high prediction performance with useful actionable interpretations. The machine learning approach is thus useful for exploring potential factors and providing new insights for fall prevention strategies.

---

**Corresponding Author:** Takaaki Ikeda, RPT, PhD; Department of Health Policy Science, Graduate School of Medical Science, Yamagata University, , YamagataYamagata, Japan (e-mail: tiked@med.id.yamagata-u.ac.jp).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11606-022-07394-8>.

**Author contribution** T. I. and U. C. are contributed equally and responsible for the conception and design, acquisition of data, analysis and interpretation of data, drafting the article, critical revision, and approval of the final manuscript. M. H., J. A., K. K., M. M., and K. O. are responsible for conception and design, interpretation of data, critical revision, and approval of the final manuscript.

**Funding** This work was supported by the Japan Society for the Promotion of Science KAKENHI (19K19818, 20H00557) and the Japan Agency for Medical Research and Development (AMED: JP17dk0110017, JP18dk0110027, JP18ts0110002, JP18le0110009, JP20dk0110034, JP20dk0110037).

**Declarations:**

**Conflict of interest:** The authors declare that they do not have a conflict of interest.

**REFERENCES**

- James SL, Lucchesi LR, Bisignano C, et al. The global burden of falls: Global, regional and national estimates of morbidity and mortality from the Global Burden of Disease Study 2017. *BMJ*. 2020;26:i3-111. doi:<https://doi.org/10.1136/injuryprev-2019-043286>
- World Health Organization. Injuries and Violence: The Fact 2014. 2014. [http://www.who.int/violence\\_injury\\_prevention/media/news/2015/Injury\\_violence\\_facts\\_2014/en/](http://www.who.int/violence_injury_prevention/media/news/2015/Injury_violence_facts_2014/en/).
- Moreland B, Kakara R, Henry A. Trends in Nonfatal Falls and Fall-Related Injuries Among Adults Aged  $\geq 65$  Years — United States, 2012–2018. *MMWR Morb Mortal Wkly Rep*. 2020;69(27):875-881. doi:<https://doi.org/10.15585/mmwr.mm6927a5>
- Florence CS, Bergen G, Atherly A, Burns E, Stevens J, Drake C. Medical Costs of Fatal and Nonfatal Falls in Older Adults. *J Am Geriatr Soc*. 2018;66(4):693-698. doi:<https://doi.org/10.1111/jgs.15304>
- Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: A systematic review and meta-analysis. *Epidemiology*. 2010;21(5):658-668. doi:<https://doi.org/10.1097/EDE.0b013e3181e89905>
- Lusardi MM, Fritz S, Middleton A, et al. Determining Risk of falls in community dwelling older adults: A systematic review and meta-analysis using posttest probability. *J Geriatr Phys Ther*. 2017;40(1):1-36. doi:<https://doi.org/10.1519/JPT.0000000000000099>
- Muir SW, Gopaul K, Montero Odasso MM. The role of cognitive impairment in fall risk among older adults: A systematic review and meta-analysis. *Age Ageing*. 2012;41(3):299-308. doi:<https://doi.org/10.1093/ageing/afs012>
- Trevisan C, Crippa A, Ek S, et al. Nutritional Status, Body Mass Index, and the Risk of Falls in Community-Dwelling Older Adults: A Systematic Review and Meta-Analysis. *J Am Med Dir Assoc*. 2019;20(5):569-582.e7. doi:<https://doi.org/10.1016/j.jamda.2018.10.027>
- Neri SGR, Oliveira JS, Dario AB, Lima RM, Tiedemann A. Does Obesity Increase the Risk and Severity of Falls in People Aged 60 Years and Older? A Systematic Review and Meta-Analysis of Observational Studies. *Journals Gerontol - Ser A Biol Sci Med Sci*. 2020;75(5):952-960. doi:<https://doi.org/10.1093/gerona/glz272>
- Soares WJS, Lopes AD, Nogueira E, Candido V, De Moraes SA, Perracini MR. Physical activity level and risk of falling in community-dwelling older adults: Systematic review and meta-analysis. *J Aging Phys Act*. 2019;27(1):34-43. doi:<https://doi.org/10.1123/japa.2017-0413>
- Seppala LJ, van de Griend EMM, Daams JG, et al. Fall-Risk-Increasing Drugs: A Systematic Review and Meta-analysis: III. Others. *J Am Med Dir Assoc*. 2018;19:372.e1-372.e8. doi:<https://doi.org/10.1016/j.jamda.2017.12.099>
- Chiarelli PE, Mackenzie LA, Osmotherly PG. Urinary incontinence is associated with an increase in falls: a systematic review. *Aust J Physiother*. 2009;55(2):89-95. doi:[https://doi.org/10.1016/S0004-9514\(09\)70038-8](https://doi.org/10.1016/S0004-9514(09)70038-8)
- Scheffer AC, Schuurmans MJ, Van dijk N, Van der hooft T, De rooij SE. Fear of falling: Measurement strategy, prevalence, risk factors and consequences among older persons. *Age Ageing*. 2008;37(1):19-24. doi:<https://doi.org/10.1093/ageing/afm169>
- Ryu E, Juhn YJ, Wheeler PH, et al. Individual housing-based socioeconomic status predicts risk of accidental falls among adults. *Ann Epidemiol*. 2017;27(7):415-420. doi:<https://doi.org/10.1016/j.annepidem.2017.05.019>
- Chang VC, Do MT. Risk factors for falls among seniors: Implications of gender. *Am J Epidemiol*. 2015;181(7):521-531. doi:<https://doi.org/10.1093/aje/kwu268>
- Hayashi T, Kondo K, Suzuki K, Yamada M, Matsumoto D. Factors associated with falls in community-dwelling older people with focus on participation in sport organizations: The Japan gerontological evaluation study project. *Biomed Res Int*. 2014;537614. doi:<https://doi.org/10.1155/2014/537614>
- Kamide N, Shiba Y, Sakamoto M, Sato H, Kawamura A. Fall-related efficacy is a useful and independent index to detect fall risk in Japanese community-dwelling older people: A 1-year longitudinal study. *BMC Geriatr*. 2019;19(1):1-9. doi:<https://doi.org/10.1186/s12877-019-1318-5>
- Racine AM, Tommet D, D'Aquila ML, et al. Machine Learning to Develop and Internally Validate a Predictive Model for Post-operative Delirium in a Prospective, Observational Clinical Cohort Study of Older Surgical Patients. *J Gen Intern Med*. 2021;36:265-273. doi:<https://doi.org/10.1007/s11606-020-06238-7>
- Klug M, Barash Y, Bechler S, et al. A Gradient Boosting Machine Learning Model for Predicting Early Mortality in the Emergency Department Triage: Devising a Nine-Point Triage Score. *J Gen Intern Med*. 2020;35:220-227. doi:<https://doi.org/10.1007/s11606-019-05512-7>
- Shakeri Hossein Abad Z, Maslove D, Lee J. Predicting Discharge Destination of Critically Ill Patients Using Machine Learning. *IEEE J Biomed Heal Informatics*. 2020:1-1. doi:<https://doi.org/10.1109/jbhi.2020.2995836>
- Lo Y, Lynch SF, Urbanowicz RJ, et al. Using machine learning on home health care assessments to predict fall risk. *Stud Health Technol Inform*. 2019;264:684-688. doi:<https://doi.org/10.3233/SHTI190310>
- Speiser JL, Callahan KE, Houston DK, et al. Machine Learning in Aging: An Example of Developing Prediction Models for Serious Fall Injury in Older Adults. *J Gerontol A Biol Sci Med Sci*. 2021;76(4):647-654. doi:<https://doi.org/10.1093/gerona/glaa138>
- Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381(9868):752-762. doi:[https://doi.org/10.1016/S0140-6736\(12\)62167-9](https://doi.org/10.1016/S0140-6736(12)62167-9)
- Chen T, Guestrin C. XGBoost: A scalable tree boosting system. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*. 2016:785-794.
- Breiman L. Random Forests. *Mach Learning*. 2001;45(1):5-32.
- Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. *Am J Epidemiol*. 2019;188(12):2222-2239. doi:<https://doi.org/10.1093/aje/kwz189>
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems*. Vol 2017 ; 2017:4766-4775.
- Lian Y, Wang GP, Chen GQ, Jia CX. Bidirectional Associations between Handgrip Strength and Depressive Symptoms: A Longitudinal Cohort Study. *J Am Med Dir Assoc*. 2021;22(8):1744-1750.e1. doi:<https://doi.org/10.1016/j.jamda.2021.04.006>
- Kondo K. Progress in Aging Epidemiology in Japan: The JAGES Project. *J Epidemiol*. 2016;26(7):331-336. doi:<https://doi.org/10.2188/jea.JE20160093>
- Ikeda T, Tsuboya T, Aida J, et al. Income and education are associated with transitions in health status among community-dwelling older people in Japan: The JAGES cohort study. *Fam Pract*. 2019;36(6):713-722. doi:<https://doi.org/10.1093/famppra/cmz022>
- Lord SR, Clark RD, Webster IW. Physiological Factors Associated with Falls in an Elderly Population. *J Am Geriatr Soc*. 1991;39(12):1194-1200. doi:<https://doi.org/10.1111/j.1532-5415.1991.tb03574.x>
- Nevitt MC, Cummings SR, Kidd S, Black D. Risk Factors for Recurrent Nonsyncopal Falls: A Prospective Study. *JAMA J Am Med Assoc*. 1989;261(18):2663-2668. doi:<https://doi.org/10.1001/jama.1989.03420180087036>
- Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112-118. doi:<https://doi.org/10.1093/bioinformatics/btr597>
- Wada T, Ishine M, Kita T, Fujisawa M, Matsubayashi K. Depression screening of elderly community-dwelling Japanese. *J Am Geriatr Soc*. 2003;51(9):1328-1329. doi:<https://doi.org/10.1046/j.1532-5415.2003.514207.x>
- Koyano W, Shibata H, Nakazato K, Haga H, Suyama Y. Measurement of competence: reliability and validity of the TMIG Index of Competence. *Arch Gerontol Geriatr*. 1991;13(2):103-116. doi:[https://doi.org/10.1016/0167-4943\(91\)90053-S](https://doi.org/10.1016/0167-4943(91)90053-S)
- Antonovsky A. *Unraveling the Mystery of Health: How People Manage Stress and Stay Well*. San Francisco: Jossey-Bass; 1987.
- Hintermair M. Sense of coherence: a relevant resource in the coping process of mothers of deaf and hard-of-hearing children? *J Deaf Stud Deaf Educ*. 2004;9(1):15-26. doi:<https://doi.org/10.1093/deaf/eh005>
- Yazawa A, Inoue Y, Fujiwara T, et al. Association between social participation and hypertension among older people in Japan: The JAGES Study. *Hypertens Res*. 2016;39(11):818-824. doi:<https://doi.org/10.1038/hr.2016.78>
- Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw*. 2010;36(11):1-13. doi:<https://doi.org/10.18637/jss.v036.i11>



40. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform.* 2019;20(2):492-503. doi:<https://doi.org/10.1093/bib/bbx124>
41. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics.* 2006;7:1-8. doi:<https://doi.org/10.1186/1471-2105-7-91>
42. Abdar M, Zomorodi-Moghadam M, Zhou X, et al. A new nested ensemble technique for automated diagnosis of breast cancer. *Pattern Recognit Lett.* 2020;132:123-131. doi:<https://doi.org/10.1016/j.patrec.2018.11.004>
43. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res.* 2012;13:281-305.
44. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explor Newsl.* 2004;6:20-29.
45. Mohammed R, Rawashdeh J, Abdullah M. Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results. *2020 11th Int Conf Inf Commun Syst ICICS 2020.* 2020:243-248. doi:<https://doi.org/10.1109/ICICS49469.2020.239556>
46. Kaur P, Gosain A. Comparing the behavior of oversampling and undersampling approach of class imbalance learning by combining class imbalance problem with noise. *Adv Intell Syst Comput.* 2018;653:23-30. doi:[https://doi.org/10.1007/978-981-10-6602-3\\_3](https://doi.org/10.1007/978-981-10-6602-3_3)
47. Aoyagi K, Ross PD, Davis JW, Wasnich RD, Hayashi T, Takemoto TI. Falls among community-dwelling elderly in Japan. *J Bone Miner Res.* 1998;13(9):1468-1474. doi:<https://doi.org/10.1359/jbmr.1998.13.9.1468>
48. Singh DKA, Shahar S, Vanoh D, Kamaruzzaman SB, Tan MP. Diabetes, arthritis, urinary incontinence, poor self-rated health, higher body mass index and lower handgrip strength are associated with falls among community-dwelling middle-aged and older adults: Pooled analyses from two cross-sectional Malaysian datas. *Geriatr Gerontol Int.* 2019;19(8):798-803. doi:<https://doi.org/10.1111/ggi.13717>
49. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. *Gen Hosp Psychiatry.* 2009;31(4):306-315. doi:<https://doi.org/10.1016/j.genhosppsych.2009.04.002>
50. Wysokiński A, Sobów T, Kloszewska I, Kostka T. Mechanisms of the anorexia of aging—a review. *Age (Omaha).* 2015;37:81. doi:<https://doi.org/10.1007/s11357-015-9821-x>
51. Forman-Hoffman VL, Yankey JW, Hillis SL, Wallace RB, Wolinsky FD. Weight and depressive symptoms in older adults: Direction of influence? *Journals Gerontol - Ser B Psychol Sci Soc Sci.* 2007;62:43-51. doi:<https://doi.org/10.1093/geronb/62.1.S43>
52. Yuki A, Otsuka R, Tange C, et al. Daily Physical Activity Predicts Frailty Development Among Community-Dwelling Older Japanese Adults. *J Am Med Dir Assoc.* 2019;20(8):1032-1036. doi:<https://doi.org/10.1016/j.jamda.2019.01.001>
53. Landi F, Russo A, Liperoti R, et al. Anorexia, physical function, and incident disability among the frail elderly population: Results from the iSIRENTE study. *J Am Med Dir Assoc.* 2010;11(4):268-274. doi:<https://doi.org/10.1016/j.jamda.2009.12.088>
54. Sairenchi T, Haruyama Y, Ishikawa Y, Wada K, Kimura K, Muto T. Sense of coherence as a predictor of onset of depression among Japanese workers: A cohort study. *BMC Public Health.* 2011;11:205. doi:<https://doi.org/10.1186/1471-2458-11-205>
55. Lakey SL, Lacroix AZ, Gray SL, et al. Antidepressant Use, Depressive Symptoms, and Incident Frailty in Women Aged 65 and Older from the Women's Health Initiative Observational Study NIH Public Access. *J Am Geriatr Soc.* 2012;60(5):854-861. doi:<https://doi.org/10.1111/j.1532-5415.2012.03940.x>
56. Varian HR. Big data: New tricks for econometrics. *J Econ Perspect.* 2014;28(2):3-28. doi:<https://doi.org/10.1257/jep.28.2.3>

**Publisher's Note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.