

Machine learning versus binomial logistic regression analysis for fall risk based on SPPB scores in older adult outpatients

DIGITAL HEALTH
Volume 9: 1–9
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231219438
journals.sagepub.com/home/dhj



Sho Hasegawa^{1,2} , Fumihiro Mizokami^{2,3}, Yoshitaka Kameya⁴,
Yuji Hayakawa^{2,5}, Tsuyoshi Watanabe⁶ and Yasumoto Matsui⁷

Abstract

Objective: To compare the performance of the diagnostic model for fall risk based on the short physical performance battery (SPPB) developed using commercial machine learning software (MLS) and binomial logistic regression analysis (BLRA).

Methods: We enrolled 797 out of 850 outpatients who visited the clinic between March 2016 and November 2021. Patients were categorized into the development ($n = 642$) and validation ($n = 155$) datasets. Age, sex, number of comorbidities, number of medications, body mass index (BMI), calf circumference (left-right average), handgrip strength (left-right average), total SPPB score, and history of falls were determined. We defined fall risk by an SPPB score of ≤ 6 in men and ≤ 9 in women. The main metrics used for evaluating the machine learning model and BLRA were the area under the curve (AUC), accuracy, precision, recall (sensitivity), specificity, and F-measure. The commercial MLS automatically calculates the parameter range of the highest contribution.

Results: The participants included 797 outpatients (mean age, 76.3 years; interquartile range, 73.0–81.0; 288 men). The metrics of the current diagnostic model in the commercial MLS were as follows: AUC = 0.78, accuracy = 0.74, precision = 0.46, recall (sensitivity) = 0.81, specificity = 0.71, F-measure = 0.59. The metrics of the current diagnostic model in the BLRA were as follows: AUC = 0.77, accuracy = 0.75, precision = 0.47, recall (sensitivity) = 0.67, specificity = 0.77, F-measure = 0.55. The risk factors for falls in older adult outpatients were handgrip strength, female sex, experience of falls, BMI, and calf circumference in the commercial MLS.

Conclusions: The diagnostic model for fall risk based on SPPB scores constructed using commercial MLS is noninferior to BLRA.

Keywords

Falls, short physical performance battery, diagnostic model, machine learning, binomial logistic regression analysis

Submission date: 14 July 2023; Acceptance date: 21 November 2023

¹Department of Pharmacotherapeutics and Informatics, Fujita Health University, Toyoake, Japan

²Department of Education and Innovation, Training for Pharmacy, National Center for Geriatrics and Gerontology, Obu, Japan

³Department of Pharmacy, National Center for Geriatrics and Gerontology, Obu, Japan

⁴Department of Information Engineering, Meijo University, Nagoya, Japan

⁵Department of Pharmacy, National Hospital Organization, Nagoya Medical Center, Nagoya, Japan

⁶Department of Orthopaedic Surgery, National Center for Geriatrics and Gerontology, Obu, Japan

⁷Center for Frailty and Locomotive Syndrome, National Center for Geriatrics and Gerontology, Obu, Japan

Corresponding author:

Sho Hasegawa, Department of Pharmacotherapeutics and Informatics, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi, 470-1192, Japan.

Email: hasegawasyo19900531@gmail.com



Introduction

More than 95% of hip fractures are caused by falls.¹ Falls in the older population are associated with high morbidity and mortality rates.² The incidence of falls increases with age, and in Japan, the annual incidence rate of falls in older adults is approximately 20%.³

In a Japanese study, a receiver operating curve analysis revealed that the consumption of five or more drugs is a significant risk factor for falls in older patients.⁴ Although polypharmacy, as an independent variable, has been linked to falls in older adults, there appears to be a stronger link between falls and the type of medications taken (e.g., medications known to increase the risk of falls) than between falls and polypharmacy itself.⁵ In older adults, falls mostly occur due to medication and multiple other causes, such as muscle weakness, age-related decline in balance, gait stability, and cardiovascular function, acute illness, mild cognitive impairment, and environmental changes.³ Therefore, the diagnosis of fall risk in older adults is difficult due to the multifactorial etiology of falls.

However, the diagnosis of fall risk is essential in older adults, especially for healthcare professionals caring for such individuals.⁶ Many diagnostic tools for fall risk have been proposed for hospitalized patients, including the Morse Fall, STRATIFY, and Hendrich Scales.⁶ Furthermore, the accuracy of fall risk assessment in older adults is reportedly comparable between conventional models based on biomechanical features and models that adopt machine learning of accelerometer data acquired in the home environment.⁷ In a Taiwanese study, researchers adopted a machine learning approach based on electronic health records with comprehensive geriatric assessment (CGA) to develop a fall risk diagnostic model for older adults.⁶ The short physical performance battery (SPPB), which is one of the CGAs, is a performance measure with clinical utility for fall risk stratification among older adults.⁸ SPPB is one of the most commonly used tools for assessing the strength of the lower extremities in older adults. It consists of three simple motor tests (tandem tests, five-timed chair stands, and gait speed measurement) that provide information on several motor domains, such as static and dynamic balance, coordination, as well as strength of the lower limbs.⁹ SPPB scores of ≤ 6 are associated with a higher fall rate in older adults of both sexes; moreover, an SPPB score of 7–9 identifies women at a higher risk of recurrent falls.⁹ However, a model for examining predictive factors for falls has not been established based on SPPB. Fall risk diagnosis using SPPB would enable effective management of older outpatients. However, the diagnosis of fall risk is limited by problems, such as increasing the workloads of healthcare professionals in terms of data collection and analysis,⁶ differences among evaluators, as well as the need for additional equipment.

Furthermore, only a few studies on diagnostic models for fall risk have focused on outpatients in Japan.^{10,11}

Logistic analysis may be useful for establishing a diagnostic model for fall risk based on SPPB scores. However, although a diagnostic model based on logistic analysis is relatively easy to use and interpret, there are some drawbacks.¹² First, the usual assumption for logistic regression analysis is that there is a linear relation between the independent and dependent variables.¹² Second, predictors are usually chosen using backward selection, which has some problems, such as being unable to reenter once a variable has been eliminated.¹² Meanwhile, machine learning has seen a rise in popularity in healthcare environments.¹³ One of the biggest distinguishing factors between binomial logistic regression analysis and machine learning is that the former emphasizes inference (i.e., infer the process of data generation), whereas the latter emphasizes prediction.¹⁴ In recent years, machine learning has become more readily available. Okawa et al. have reported that Prediction One, a commercial machine-learning software, is useful for the establishment of a diagnostic model for cisplatin-induced acute kidney injury.¹⁵ Prediction One can build an ensemble model consisting of neural networks and gradient-boosting decision trees using all predictors. Meanwhile, there is deep learning, which is a multilayered neural network in machine learning. Furthermore, there are various methods, such as XGBoost, as a framework for gradient-boosting decision trees. However, Prediction One cannot select these detailed methods, although it is easy to use. In the present study, we employed Prediction One to assess the difference in the diagnostic performance between user-friendly commercial machine learning software and statistical methods. The primary aim of this present study was to compare the predictive performance of the diagnostic model for fall risk based on SPPB developed using commercial machine learning software (Prediction One) and binomial logistic regression analysis.

Methods

Definition of participants

All patient details were de-identified in this cross-sectional study. Patients who visited the frailty clinic at the National Center for Geriatrics and Gerontology between March 2016 and November 2021 were included during their first visit. Patients lacking data were excluded. To establish a machine learning-based diagnostic model, the patients were classified into the development and validation datasets with no differences between each dataset in clinical characteristics. All participants provided written informed consent for routine medical care and registration in this study. This study was approved by the ethics review board of the National Center for Geriatrics and Gerontology

(approval no.: 881–13; 2 February 2022). This study was conducted in accordance with the World Medical Association Declaration of Helsinki. Furthermore, the study was conducted in accordance with the Strengthening of Reporting of Observational Studies in Epidemiology guidelines.¹⁶

SPPB scores for fall risk

SPPB includes tandem score (side-by-side, semi-tandem, and full tandem balance for up to 10 s each), five-timed chair stands score (ability and time required to stand five times as quickly as possible with arms folded across the chest from a straight-backed chair), and gait speed measurement score (timed 4-m walk at a self-selected pace).¹⁷ SPPB tandem score (range, 0–4), SPPB five-timed chair stands score (range, 0–4), SPPB gait speed measurement score (range, 0–4), and total SPPB score (range, 0–12) were assessed by physical therapists.¹⁸ Higher total SPPB scores indicate better physical performance.¹⁸ The SPPB scores for fall risk were obtained from a previous study⁹ that revealed an association between SPPB scores of ≤ 6 and a higher fall rate among older adults of both sexes. In women, an SPPB score of 7–9 identified those at a higher risk of recurrent falls. Thus, fall risk was defined by an SPPB score of ≤ 6 in men and ≤ 9 in women.

Measurement and data collection

Data on age, sex, number of comorbidities, medications taken, body mass index (BMI), calf circumference (left–right average), handgrip strength (left–right average), total SPPB score, and history of falls in the past 1 year were collected at the first visit. We defined fall risk based on SPPB score = 1, sex (male) = 1, and history of falls (experience of falls) = 1 in the machine learning and binomial logistic regression analysis. Medications identified as risk factors for falls were measured in accordance with the Guidance on Appropriate Medication for Elderly Patients proposed by Japan’s Ministry of Health, Labor and Welfare,¹⁹ including antihypertensive drugs, sleeping pills, antianxiety drugs, antidepressant drugs, antiepileptic drugs, antipsychotic drugs, drugs for Parkinson’s disease, antihistamines (including H₂ receptor antagonists), and memantine (eTable 1).

Machine learning methods

Machine learning was performed as described previously¹⁵ using Prediction One version 3.0.1.3 (Sony Corporation, Tokyo, Japan). In Prediction One, all predictors were used to build an ensemble model consisting of neural networks and gradient-boosting decision trees. Neural networks can learn representations of both key factors and their interactions from data while gradient-boosting decision trees

support decision-making, similar to a flowchart estimating the risk of an outcome. Prediction One automatically determines the hyperparameters of models to improve the models’ predictive performance using validation datasets. Moreover, Prediction One can identify the rank of contribution as a predictive factor based on the analysis of sensitivity using development data. Prediction One automatically calculates the range of parameters with the highest contribution using development data, as well as accuracy, precision, recall (sensitivity), and specificity at the maximum value of F-measure by using validation data. The main metrics used for evaluating the machine learning model was the area under the curve (AUC), accuracy, precision, recall (sensitivity), specificity, and F-measure.

Statistical analyses

Variables were expressed as median with interquartile range (IQR) or as prevalence for each data set. To determine the differences between the development and validation datasets with respect to participants’ characteristics at the first visit, the Wilcoxon rank-sum test was performed for continuous variables after checking for nonnormal distribution using D’Agostino’s K-squared test. Additionally, a chi-square test was performed to compare categorical variables. Contributions as a predictive factor in development data were calculated by binomial logistic regression analysis. The AUC, accuracy, precision, recall (sensitivity), and specificity at the maximum value of F-measure in the validation data (eTable 1) were calculated using regression coefficients. The above statistical analysis was performed using BellCurve for Excel (Social Survey Research Information Co., Ltd, Tokyo, Japan), and statistical significance was set at $p < 0.05$.

Results

Participant enrollment

This study included 797 of the 850 outpatients who visited the frailty clinic at the National Center for Geriatrics and Gerontology between March 2016 and November 2021. Participants were categorized into the development ($n = 642$) and validation ($n = 155$) datasets (Figure 1), with a median age (IQR) of 76.3 (73.0–81.0) years (Figure 1).

Characteristics of participants at the first visit

At the first visit, there were no significant differences between the development and validation datasets in terms of age, sex, number of comorbidities, number of medications, number of medications as risk factors for falls, BMI, calf circumference, handgrip strength, total SPPB scores, and experience of fall (Table 1). The top 10 most common comorbidities are also shown in eTable 2.

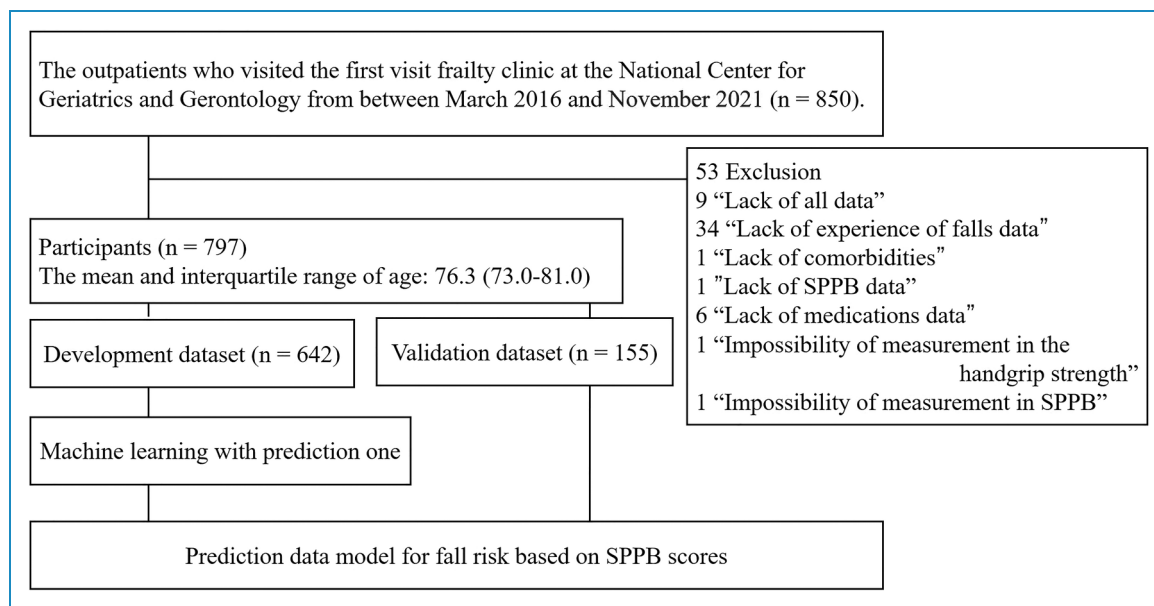


Figure 1. Enrollment of the participants.
SPPB: short physical performance battery.

Table 1. Characteristics of participants at the first visit.

| | Development (n = 642) | Validation (n = 155) | p-value |
|---|-----------------------|----------------------|---------|
| Age (years old), median (IQR) | 77.0 (73.0–81.0) | 77.0 (73.0–81) | 0.94* |
| Sex | | | |
| Male (n) | 238 | 50 | 0.26# |
| Female (n) | 404 | 105 | |
| Number of comorbidities, median (IQR) | 4.0 (2.0–5.0) | 4.0 (2.0–5.0) | 0.95* |
| Number of medications, median (IQR) | 5.0 (3.0–7.0) | 5.0 (3.0–8.0) | 0.80* |
| Number of medications as risk factors for falls, median (IQR) | 1.0 (0.0–2.75) | 2.0 (0.5–3.0) | 0.29* |
| BMI (kg/m ²), median (IQR) | 23.6 (21.3–26.3) | 23.5 (21.2–25.6) | 0.50* |
| Calf circumference (cm), median (IQR) | 34.3 (31.8–36.5) | 33.5 (32.0–35.5) | 0.08* |
| Handgrip strength (kg), median (IQR) | 22.8 (18.0–28.6) | 22.0 (17.7–26.1) | 0.14* |
| Total SPBB scores, median (IQR) | 11.0 (9.0–12.0) | 11.0 (9.0–12.0) | 0.35* |
| History of falls | | | |
| Falls (n) | 207 | 49 | 0.88# |
| Nonfalls (n) | 435 | 106 | |

*Wilcoxon rank-sum test (vs validation dataset), #Chi-square test. BMI: body mass index, SPPB: short physical performance battery.

Prediction One vs. binomial logistic regression analysis in the model performance of SPPB scores for fall risk using validation data

The receiver operating characteristic curve in Prediction One is shown in Figure 2A. The metrics of the current diagnostic model in Prediction One were as follows: AUC = 0.78, accuracy = 0.74, precision = 0.46, recall (sensitivity) = 0.81, specificity = 0.71, and F-measure = 0.59 (Table 2). Table 3 lists the rank of contribution as a predictive factor based on SPPB scores for fall risk. Handgrip strength, sex, history of falls, BMI, and calf circumference contributed the most to fall risk based on SPPB scores (Table 3). The receiver operating characteristic curve in the binomial logistic

regression analysis is shown in Figure 2B. The metrics of the current diagnostic model in the binomial logistic regression analysis were as follows: AUC = 0.77, accuracy = 0.75, precision = 0.47, recall (sensitivity) = 0.67, specificity = 0.77, and F-measure = 0.55 (Table 2). Handgrip strength, sex, history of falls, BMI, and calf circumference were significantly associated with greater fall risk based on SPPB scores (Table 3). Table 4 presents the range of parameters with the highest contribution as a predictive factor based on SPPB scores for fall risk related to each parameter in Prediction One. Handgrip strength (6.75–14.4 kg), female sex, experience of falls, BMI (27.5–30.0 kg/m²), and calf circumference (21.4–30.0 cm) were risk factors associated with fall risk (Table 4).

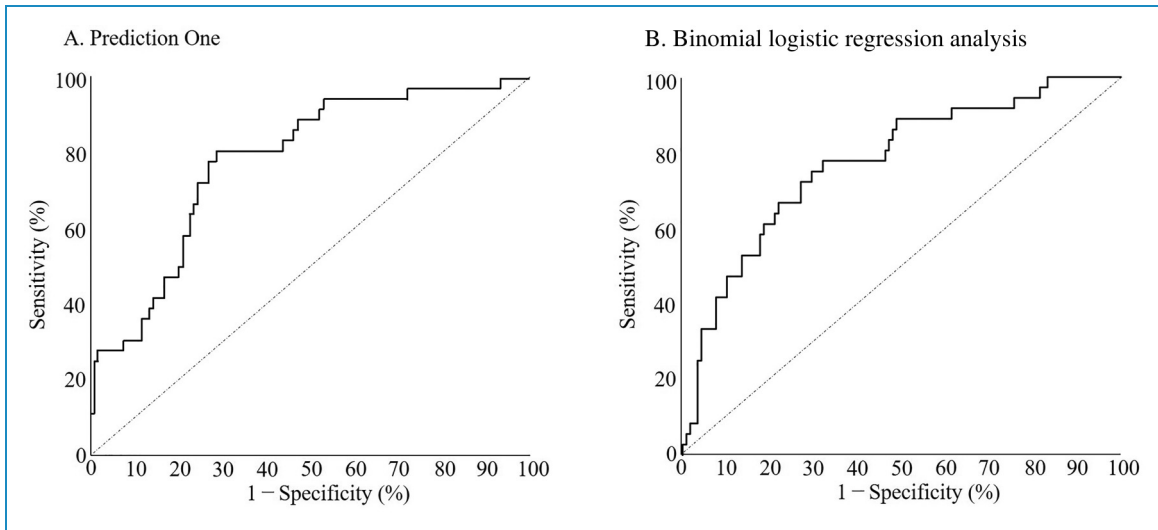


Figure 2. Model performance for diagnostic model for fall risk based on SPPB scores. Receiver operating characteristic curve of (A) Prediction One and (B) Binomial logistic regression analysis. SPPB: short physical performance battery.

Table 2. Prediction One versus binomial logistic regression analysis in the model performance of SPPB scores for falls risk using validation data.

| | | Falls risk (actual measurements) | | | | | | | | |
|------------------------------|-------------------------------|----------------------------------|---------|-----|----------|-----------|----------------------|-------------|-----------|------|
| | | Risk | Nonrisk | AUC | Accuracy | Precision | Sensitivity (Recall) | Specificity | F-measure | |
| Prediction One | Falls risk (diagnostic model) | Risk | 29 | 34 | 0.78 | 0.74 | 0.46 | 0.81 | 0.71 | 0.59 |
| | | Nonrisk | 7 | 85 | | | | | | |
| Binomial logistic regression | Falls risk (diagnostic model) | Risk | 24 | 27 | 0.77 | 0.75 | 0.47 | 0.67 | 0.77 | 0.55 |
| | | Nonrisk | 12 | 92 | | | | | | |

SPPB: short physical performance battery, AUC: area under the curve.

Table 3. Contribution as a diagnostic model for fall risk based on SPPB scores using the Prediction One and Binomial logistic regression analysis in the development data.

| Model | Age | Sex | Number of comorbidities | Number of medications | Number of medications for fall risk factors | BMI | Calf circumference | Handgrip strength | History of falls |
|---------------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|---|------------------------------|------------------------------|------------------------------|------------------------------|
| Prediction One (Rank) | 7 | 2 | 6 | 8 | 9 | 4 | 5 | 1 | 3 |
| Binomial logistic regression analysis | <i>p</i> = 0.90 [†] | <i>p</i> < 0.01 [†] | <i>p</i> = 0.38 [†] | <i>p</i> = 0.17 [†] | <i>p</i> = 0.99 [†] | <i>p</i> < 0.01 [†] | <i>p</i> < 0.01 [†] | <i>p</i> < 0.01 [†] | <i>p</i> < 0.01 [†] |
| Odds ratio | 1.00 | 0.25 | 1.05 | 1.06 | 1.00 | 1.13 | 0.86 | 0.91 | 2.31 |
| 95% CI | 0.97–1.03 | 0.13–0.48 | 0.94–1.18 | 0.97–1.16 | 0.84–1.19 | 1.05–1.23 | 0.77–0.95 | 0.87–0.95 | 1.52–3.50 |

SPPB: short physical performance battery, BMI: body mass index. [†]Binomial logistic regression analysis. We defined fall risk based on SPPB = 1, sex (male) = 1, and history of falls (experience of falls) = 1 in the machine learning and binomial logistic regression analysis.

Table 4. The parameter of highest contribution as a diagnostic model for fall risk based on SPPB scores using Prediction One in the development data.

| Characteristics | Parameter |
|--------------------------|---------------------|
| Handgrip strength (kg) | 6.75–14.4 |
| Sex | Female |
| History of falls | Experience of falls |
| BMI (kg/m ²) | 27.5–30.0 |
| Calf circumference (cm) | 21.4–30.0 |

SPPB: short physical performance battery; BMI: body mass index.

Discussion

The metrics of the current diagnostic model in the binomial logistic regression analysis were as follows: AUC = 0.77, accuracy = 0.75, precision = 0.47, recall (sensitivity) = 0.67, specificity = 0.77, and F-measure = 0.55. Meanwhile, Prediction One showed a similar performance to binomial logistic regression analysis, with AUC = 0.78, accuracy = 0.74, precision = 0.46, recall (sensitivity) = 0.82, specificity = 0.71, and F-measure = 0.59. In this study, the precision of the diagnostic model in Prediction One was low (0.46), which is expected to result in many false positives. However, a lower number of false negatives is required to prevent falls. Moreover, the recall was 0.81, suggesting that the number of false negatives was low. In a preliminary study of binomial logistic regression analysis, a recall of 0.81 was found when the F-value was 0.48 (data not shown). This F-value in the logistic analysis was also similar to the F-value in machine learning (0.59). These findings suggest that a diagnostic model for fall risk based on SPPB scores in Prediction One is noninferior to binomial logistic regression analysis. In the preliminary study, in the binary logistic regression analysis, the AUC (0.78) in the development dataset was similar to that in the validation dataset (data not shown). Meanwhile, in Prediction One, the AUC (0.93) in the development dataset was higher than that in the validation dataset (data not shown). Therefore, although the diagnostic model based on Prediction One tends to overfit, there is a possibility that the performance of the diagnostic model will improve when the sample size is increased. In the future, studies with more participants are required to achieve better precision in Prediction One.

In the present study, we established a fall risk diagnostic model based on the SPPB score according to a previous study.⁹ In that study, the mean age of participants was 75.1 years for men and 75.8 years for women; however, in the current study, the mean age was 77 years. The mean age of participants in the present study is consistent

with that in the previous study,⁹ and the setting of a cut-off value for fall risk based on the SPPB score is reasonable. Meanwhile, the cut-off value for fall risk based on the SPPB score was an index of recurrent falls. In the present study, the number of participants who had multiple experiences of falls was small, making it difficult to build a predictive model focusing on recurrent falls. Therefore, we aimed to build a model that could identify patients potentially at risk of recurrent falls and promote vigilance. Further studies with a larger sample size are required to establish a fall risk diagnostic model based on the SPPB score focusing on recurrent falls.

Toba et al. reported on the Fall Risk Index (FRI), which is a self-administered questionnaire comprising 22 questions with yes/no responses and a question on the number of falls over the past 12 months.^{10,11} The Instrumental Activities of Daily Living Scale (IADL) score, Activities of Daily Living (ADL), age, and systolic pressure are important factors that must be assessed by a diagnostic model developed using machine learning for falls.⁶ It may be difficult for participants with cognitive dysfunction to self-assess the FRI, ADL, and IADL scores using questionnaires.^{11,20,21} Therefore, objective diagnostic systems for fall risk are required to prevent the occurrence of falls. The SPPB has been used extensively to assess the physical and functional health of community-dwelling older adults.²² Poor performance on the SPPB is associated with adverse health outcomes, such as nursing home placement, increased need for caregiver support, functional decline, and mortality.²² A Taiwanese study revealed that electronic health records with CGA were useful for the construction of a diagnostic model for fall risk using machine learning.⁶ However, it is often difficult to implement electronic health records and CGA with the SPPB in the diagnosis of fall risk in older adult outpatients. Thus, we aimed to establish a simple diagnostic model for SPPB to prevent falls based on a previous study,⁹ with simple factors, such as sex, age, and handgrip strength. Our diagnostic model in Prediction One indicated that handgrip strength (6.75–14.4 kg), female sex, experience of falls, BMI (27.5–30.0 kg/m²), and calf circumference (21.4–30.0 cm) were risk factors for falls based on SPPB scores, which were consistent with the results of binomial logistic regression analysis. Furthermore, these findings are also mostly in line with those of previous studies, which revealed handgrip strength, female sex, experience of falls, and BMI as predictors of falls.^{23–25} Sarcopenia is positively correlated with falls.²⁵ In a Japanese study of 526 participants, calf circumference was positively correlated with appendicular skeletal muscle mass and skeletal muscle index and was associated with the diagnosis of sarcopenia.²⁶ Therefore, these factors may be useful for the simple and objective prevention of falls in medical institutions, such as community pharmacies, and even outside of hospitals. Further studies are required to determine if

these risk factors are useful for predicting falls, with a focus on participants with cognitive dysfunction.

The contribution of the number of medications as a risk factor for falls was low in each diagnostic model. One reason for this may be changes in the prescription tendency. The prescription of benzodiazepine receptor agonists, benzodiazepines, and nonbenzodiazepines, which are risk factors for falls, has decreased in patients aged ≥ 75 years in Japan.²⁷ The use of potentially inappropriate medications, as defined by the 2003 Beers criteria, decreased between 1988 and 2010.²⁸ Therefore, due to the decreased tendency in prescription, the number of medications as a risk factor for falls was associated with a low contribution in this diagnostic model. Further investigation is required to validate whether drugs other than those previously reported to be associated with a fall risk are truly involved in falls because orexin receptor antagonists (which tend to be replaced by benzodiazepines) are associated with the same level of fall risk as nonbenzodiazepines.²⁹

Previous research has proposed a fall risk diagnostic model using logistic regression analysis. Although their target patients and study plans were not consistent with our study, the AUCs of a fall risk diagnostic model based on logistic regression analysis using factors, including handgrip strength, were reported to be 0.79³⁰ and 0.71.³¹ It may be noninferior to previous models that also diagnostic models for fall risk because the AUC of the machine learning model in our study was 0.78. Meanwhile, the limitation of our study is that we did not build a model that directly predicted falls. The AUC of the diagnostic model for fall risk based on SPPB scores using binomial logistic regression analysis was found to be between 0.6 and 0.7.^{32–34} Therefore, it is necessary to construct a model that directly predicts falls. Prediction One has a function of time-series forecasting, and participants in the present study are being followed long-term. The World Health Organization (WHO) has categorized fall risk into four dimensions: biological (such as physical decline), behavioral (such as multiple medication use), environmental (such as slippery floors and stairs), and socioeconomic (such as lack of community resources) factors.³⁵ Based on these four dimensions, WHO also recommends the following for a comprehensive assessment of fall risk: history of falls; gait, balance, mobility, and muscle weakness; osteoporosis risk; fear of falling, vision impairment, cognitive impairment, and neurological examination; urinary incontinence; home hazards; cardiovascular examination; and medication review.³⁶ In the present study, the fall diagnostic model focused on some of the biological factors among these four dimensions. Therefore, we plan to develop this research and conduct further studies on building a model that predicts future falls, focusing on other dimensions, using Prediction One.

Furthermore, in the dataset used in this study, we were only able to ascertain a history of falls in the past 1 year.

Therefore, the cut-off for SPPB scores was based on a previous report on recurrent falls,⁹ with no clarity on the first fall. Further studies are required to establish a diagnostic model based on the history of falls. Although systolic pressure has been reported to be a useful predictor of falls,⁶ it was difficult to examine as a factor because resting blood pressure was not measured in our study.¹⁸ Thus, further investigations involving the measurement of resting blood pressure are necessary.¹⁸ The diagnostic criteria for sarcopenia,³⁷ which is positively correlated with falls,²⁵ include sex-based reference values for handgrip strength (male: 28 kg, female: 18 kg) and calf circumference (male: 34 cm, female: 33 cm). In a preliminary machine learning study, we observed no differences in predictive performance parameters, such as AUC and recall, when adding risk factors based on sex-specific sarcopenia cut-off values for handgrip strength and calf circumference (data not shown). It was difficult to predict falls for men and women separately due to the small sample sizes. Further studies involving larger sample sizes are required to establish a sex-based diagnostic model for falls.

Conclusions

The diagnostic model for fall risk based on SPPB scores constructed using commercial machine learning software is noninferior to binomial logistic regression analysis. We plan to conduct further studies on building a model that predicts future falls using this software.

Acknowledgments: We would like to thank Editage (<https://www.editage.com/>) for editing and reviewing this manuscript for the English language.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: This work was supported by the Research Funding for Longevity Sciences (grant number 20-42) from the National Center for Geriatrics and Gerontology (NCGG), Japan; a JSPS Grant-in-Aid for Early-Career Scientists (grant number 22K15352); and MHLW Research on Dementia Program (grant number 21GB1002).

Ethical approval: The ethics review board of the National Center for Geriatrics and Gerontology (approval no.: 881-13; 2 February 2022) approved this study.

Guarantor: SH.

Contributorship: SH, FM, YK, and TW conceptualized this paper. YM and TW were involved in gaining ethical approval and patient recruitment. SH and FM organized the patient data. SH, FM, and YK were involved in the design of the methodology and analyzed the data. SH, FM, YK, and YH

interpreted the results. SH and FM wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

ORCID ID: Sho Hasegawa  <https://orcid.org/0000-0003-1236-1623>

Supplemental material: Supplemental material for this article is available online.

References

1. Vaishya R and Vaish A. Falls in older adults are serious. *Indian J Orthop* 2020; 54: 69–74.
2. de Jong MR, Van der Elst M and Hartholt KA. Drug-related falls in older patients: implicated drugs, consequences, and possible prevention strategies. *Ther Adv Drug Saf* 2013; 4: 147–154.
3. Nakanishi T, Ikeda T, Nakamura T, et al. Development of an algorithm for assessing fall risk in a Japanese inpatient population. *Sci Rep* 2021; 11: 17993.
4. Kojima T, Akishita M, Nakamura T, et al. Polypharmacy as a risk for fall occurrence in geriatric outpatients. *Geriatr Gerontol Int* 2012; 12: 425–430.
5. Hammond T and Wilson A. Polypharmacy and falls in the elderly: a literature review. *Nurs Midwifery Stud* 2013; 2: 171–175.
6. Chu WM, Kristiani E, Wang YC, 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 (Lausanne)* 2022; 9: 937216.
7. Nait Aicha A, Englebienne G, van Schooten KS, et al. Deep learning to predict falls in older adults based on daily-life trunk accelerometry. *Sensors (Basel)* 2018; 18: 1654.
8. Welch SA, Ward RE, Beauchamp MK, et al. The Short Physical Performance Battery (SPPB): a quick and useful tool for fall risk stratification among older primary care patients. *J Am Med Dir Assoc* 2021; 22: 1646–1651.
9. Veronese N, Bolzetta F, Toffanello ED, et al. Association between short physical performance battery and falls in older people: the Progetto Veneto Anziani study. *Rejuvenation Res* 2014; 17: 276–284.
10. Toba K, Okochi J, Takahashi T, et al. Development of a portable fall risk index for elderly people living in the community. *Nihon Ronen Igakkai Zasshi* 2005; 42: 346–352.
11. Kikuchi R, Kozaki K, Iwata A, et al. Evaluation of risk of falls in patients at a memory impairment outpatient clinic. *Geriatr Gerontol Int* 2009; 9: 298–303.
12. Ingwersen EW, Stam WT, Meijs BJV, et al. Machine learning versus logistic regression for the prediction of complications after pancreateoduodenectomy. *Surgery* 2023; 174: 435–440.
13. Song X, Liu X, Liu F, et al. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *Int J Med Inform* 2021; 151: 104484.
14. Liew BXW, Kovacs FM, Rugamer D, et al. Machine learning versus logistic regression for prognostic modelling in

- individuals with non-specific neck pain. *Eur Spine J* 2022; 31: 2082–2091.
15. Okawa T, Mizuno T, Hanabusa S, et al. Prediction model of acute kidney injury induced by cisplatin in older adults using a machine learning algorithm. *PLoS One* 2022; 17: e0262021.
 16. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577.
 17. Riskowski JL, Hagedorn TJ, Dufour AB, et al. Functional foot symmetry and its relation to lower extremity physical performance in older adults: the Framingham Foot Study. *J Biomech* 2012; 45: 1796–1802.
 18. Hasegawa S, Mizokami F, Mase H, et al. Effects of discontinuation of antihypertensive drugs on frailty syndrome in outpatients: a 1-year prospectively designed retrospective chart-review pilot study. *J Int Med Res* 2022; 50: 300060522 1130716.
 19. Guidance on Appropriate Medication for Elderly Patients proposed by Japan's Ministry of Health, Labor and Welfare, <https://www.pmda.go.jp/files/000232249.pdf> (2018, accessed 1 October 2022 [In Japanese])
 20. Lawton MP and Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969; 9: 179–186.
 21. Silveira LTYD, Silva JMD, Soler JMP, et al. Assessing functional status after intensive care unit stay: the Barthel Index and the Katz Index. *Int J Qual Health Care* 2018; 30: 265–270.
 22. Fisher S, Ottenbacher KJ, Goodwin JS, et al. Short physical performance battery in hospitalized older adults. *Aging Clin Exp Res* 2009; 21: 445–452.
 23. Gale CR, Cooper C and Aihie Sayer A. Prevalence and risk factors for falls in older men and women: the English Longitudinal Study of Ageing. *Age Ageing* 2016; 45: 789–794.
 24. Villamizar-Pita PC, Angarita-Fonseca A, de Souza HCD, et al. Handgrip strength is associated with risk of falls in physically active older women. *Health Care Women Int* 2022; 43: 1301–1314.
 25. Yeung SSY, Reijnierse EM, Pham VK, et al. Sarcopenia and its association with falls and fractures in older adults: a systematic review and meta-analysis. *J Cachexia Sarcopenia Muscle* 2019; 10: 485–500.
 26. Sun YS, Kao TW, Chang YW, et al. Calf circumference as a novel tool for risk of disability of the elderly population. *Sci Rep* 2017; 7: 16359.
 27. Okui T, Park J, Hirata A, et al. Trends in the prescription of benzodiazepine receptor agonists from 2009 to 2020: a retrospective study using electronic healthcare record data of a university hospital in Japan. *Healthcare (Basel)* 2021; 9: 1724.
 28. Charlesworth CJ, Smit E, Lee DS, et al. Polypharmacy among adults aged 65 years and older in the United States: 1988–2010. *J Gerontol A Biol Sci Med Sci* 2015; 70: 989–995.
 29. Ishibashi Y, Nishitani R, Shimura A, et al. Non-GABA sleep medications, suvorexant as risk factors for falls: case-control and case-crossover study. *PLoS One* 2020; 15: e0238723.
 30. Stalenhoef PA, Diederiks JP, Knottnerus JA, et al. A risk model for the prediction of recurrent falls in community-dwelling elderly: a prospective cohort study. *J Clin Epidemiol* 2002; 55: 1088–1094.
 31. Nagai T, Okano I, Ishikawa K, et al. The serum 25(OH)D level and hand grip strength for fall risk assessment among osteoporotic elderly Japanese women. *Arch Osteoporos* 2021; 16: 42.
 32. Li W, Rao Z, Fu Y, et al. Value of the short physical performance battery (SPPB) in predicting fall and fall-induced injury among old Chinese adults. *BMC Geriatr* 2023; 23: 574.
 33. Fukui K, Maeda N, Komiya M, et al. The relationship between modified short physical performance battery and falls: a cross-sectional study of older outpatients. *Geriatrics (Basel)* 2021; 6: 106.
 34. Lauretani F, Ticinesi A, Gionti L, et al. Short-Physical Performance Battery (SPPB) score is associated with falls in older outpatients. *Aging Clin Exp Res* 2019; 31: 1435–1442.
 35. WHO global report on falls prevention in older age. <https://www.who.int/publications/i/item/9789241563536> (2008, accessed 11 October 2023).
 36. Integrated care for older people: guidelines on community-level interventions to manage declines in intrinsic capacity. <https://www.who.int/publications/i/item/9789241550109> (2017, accessed 11 October 2023).
 37. Chen LK, Woo J, Assantachai P, et al. Asian working group for sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. *J Am Med Dir Assoc* 2020; 21: 300–307 e2.
-