Identifying key fall risk factors in older adults within a multifactorial intervention using feature selection methods

Raul R. Teles[1], Mariana L. de Melo[2], Karina Gramani-Say[2], Juliana H. Ansai[2], Paulo H. S. Pelicioni[3], Jean R. Ponciano[1]

[1]Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP)
São Carlos – SP – Brazil
raulteles@usp.br, jeanponciano@icmc.usp.br

[2]Departamento de Gerontologia – Universidade Federal de São Carlos (UFSCar)
São Carlos – SP – Brazil
marianamelo@estudante.ufscar.br, {gramanisay,jhansai}@ufscar.br

[3]Universidade de New South Wales (UNSW) – New South Wales, SYD – Australia
paulo.silvapelicioni@unsw.edu.au

Abstract. *Approximately 27% of older adults experience fall-related injuries each year. To reduce falls, multifactorial interventions assess individual risk factors and develop targeted care plans. While these interventions may reduce falls, their success depends on the accurate identification of risk factors and the implementation of appropriate plans. In this paper, we applied feature selection methods to identify relevant variables for predicting falls in three distinct moments of an intervention, achieving precisions of 0.6, 0.6 and 0.57 and recalls 0.71 0.75 and 1.0. Relevant factors found include mobility, anxiety and depression, fear of falling, home hazards, physical activity, and income. We expect these results to assist domain experts by enabling data-driven decision-making.*

1. Introduction

Given the world's increasing life expectancy, especially in developing countries such as Brazil, age-related health concerns are becoming increasingly relevant. One particularly important concern is fall-related injuries, the second biggest cause of unintentional injury in the world, affecting approximately 27% of older adults each year [Saunders et al. 2025]. Falls may result not only in physical injuries, such as fractures and hospitalizations, but also in depression, functional decline, and even in societal and economic problems, including expenses in public health systems, social isolation, and institutionalization [Li et al. 2023]. In the worst cases, falls can lead to death.

In this context, multifactorial interventions are programs that aim to reduce falls in older adults by tailoring multiple fall-prevention strategies, such as exercise and medical review, to each patient's individual risk factors [Hopewell et al. 2018]. They can be delivered via case management programs, in which case managers tailor interventions according to the older people's preferences and risks, and monitor these patients to identify and address any difficulties in following the care plans proposed along the program [Alberto et al. 2022]. These plans may include, for example, exercise routines focusing on mobility, strength, and balance training, periodic evaluations on lower limbs' strength and mobility, general health tips, and strategies for reducing home hazards [Alberto et al. 2022]. While these tailor-made plans have the potential to engage patients and address difficulties that would not be found in non-personalized plans, their success depends on the correct identification of the individual's risk factors, which is a challenge considering time and resource constraints, such as the number of patients per clinician, the size of the staff, and the budget for sophisticated equipment (e.g., sensors). Moreover, the scientific community still does not fully understand the influence of certain risk factors, such as psychosocial and demographic factors [Saunders et al. 2025], making it harder to identify priorities for care plans.

In this paper, we applied computational feature selection methods in data coming from a real-world case management multifactorial intervention to identify features – that is, demographic, behavioral, and physiological factors – that better predict the occurrence of falls in a 6-month period after the program ended. We thus aim to provide an initial step toward understanding which factors should be prioritized under limited intervention time and resource constraints, thereby supporting clinicians and case managers in more informed decision-making. Another contribution is that this study, as far as the authors know, is the first to apply and evaluate machine learning techniques on intervention data, and to compare the results and selected features across different stages of the intervention.

2. Background and related works

2.1. The MAGIC study

Our study originates from the *Case Management Program at Home to Reduce Fall Risk in Older Adults (MAGIC)*, conducted by *Grupo de Pesquisa Abordagem funcional e multiprofissional em Gerontologia* (Gerontology Department/UFSCar) with community-dwelling older adults from São Carlos - SP, Brazil [Alberto et al. 2022]. The participants were adults over 60 years old who had experienced at least two falls in the year prior to the program. They were separated into intervention and control groups, with the intervention group undergoing a multidimensional fall risk assessment, having their risk factors explained and partaking an intervention program devised based on their identified factors and preferences. The intervention spanned 16 weeks, and the case manager helped the patients with any difficulties in executing their care plans.

Multifactorial assessments were collected at the beginning of the intervention (baseline), at the end of the 16 weeks, and 6 weeks after the intervention ended. Then, there was a 1-year follow-up. The assessments included sociodemographic measures, body mass index, medication use, morbidities, foot problems, history of falls, hospitalizations, physical activity, functional activity, and fracture risk [Alberto et al. 2022]. Prospective fall data were collected across all intervention stages, including the follow up.

2.2. Predicting falls with machine learning

Machine learning algorithms have been used in a wide range of fall contexts, including detecting falls as soon as they occur to minimize consequences, estimating fall risk using postural stability [Liang et al. 2024], and predicting falls using sensors [Maruf et al. 2025]. Training a machine learning model to predict falls is a specific case of a problem known as classification, in which the model receives a sample and predicts to which class it is part of, e.g., determining if a skin disease image indicates a cancer diagnosis, or, in the case of falls, if a patient will fall in the next 6 months or won't. A literature review from 2025 [Capodici et al. 2025], that focused exclusively on studies on fall prediction with no sensor data, shows that models are usually trained on datasets of different populations (inpatients, community dwellers, populations with certain health conditions), sample sizes (42 to almost 300,000), and using different feature sets. Regardless, many studies achieved good results using different models tailored to their specific data, which highlights the flexibility of machine learning techniques and the importance of studying them in specific contexts.

2.3. Recursive Feature Elimination

There are many computational techniques for feature selection. Among them, Recursive Feature Elimination (RFE) is a method that works by repeatedly training a classifier algorithm that can rank the features it was fitted in. It consists of the following three-step iterative procedure [Guyon et al. 2002]: (i) train the classifier; (ii) compute the ranking criterion for all features; (iii) remove the feature with the smallest ranking criterion. This loop can then end when a target number of features is achieved.

To optimize the number of features, RFE can be trained within different cross-validation splits, calculating scores for different numbers of features. That makes it possible to take the number of features that have the highest average score across splits.

2.4. Feature selection on health data

Feature selection is widely used on health data as it is more interpretable when compared to feature extraction, since it maintains the context and real-world meaning of the original features. Areas of application of feature selection include medical imaging, biomedical signal processing, and DNA data [Remeseiro and Bolon-Canedo 2019]. Closest to our work, feature selection has also been used on some fall prediction models to improve accuracy [Liang et al. 2024][Zhang et al. 2025][Ebrahimi et al. 2022].

3. Methodology

3.1. Preparing the dataset

The data comes from the MAGIC study (Section 2.1). It contains the multifactorial assessments from the baseline, the end of the intervention, and 6 weeks after it finished. For convenience, we will refer to these moments in time as weeks 0, 16, and 22. The dataset has data of 62 patients (samples), with 182, 152, and 152 features for the weeks in order. The number-of-falls variable used in our study comes from the 1-year follow-up.

The dataset was preprocessed before starting the analysis and feature selection. First, people who left the intervention before the end were removed, along with those who had more than 20% of their values missing. Features with more than 80% null values were also removed. After removing these samples and features, data inconsistencies were manually fixed, such as numerical features containing annotations and out-of-range values, which were replaced with the closest possible value. Non-numerical features and features with no variance were removed, and duplicated features were retained only once. Then, categorical features with more than two classes were discarded, except for race and marital status, which were one-hot encoded, and features that were components of more general features were discarded. Lastly, we added the features that were collected only at baseline, mainly socioeconomic ones (e.g., age, income, sex, race), to weeks 16 and 22.

Table 1. Evaluated parameter combinations

Model	Parameters
LR	$C \in \{0.01,\ 0.03,\ 0.10,\ 0.32,\ \underline{1.0}\}$
SVM	$C \in \{0.001,\ 0.002,\ 0.005,\ 0.010,\ 0.022,\ \underline{0.046},\ 0.100,\ 0.215,\ 0.464,\ 1.000\}$ penalty $\in \{$`'l1'`, <u>`'l2'`</u>$\}$
RF	max_features $\in \{$<u>`sqrt`</u>, `log2`$\}$ max_depth $\in \{$`None`, 5, <u>10</u>$\}$

The final dataset ended up with 50 patients (21 fallers, 29 non-fallers) and 78, 84, and 87 numerical features (at least encoded) for the measured weeks, in order. Some of the remaining features involved weight, height, socioeconomic factors (income, years of schooling), demographics factors (age, race, sex), number of morbidities and medications, pain assessments, general falls assessment questionnaires, fear of falling, cognition assessments, functional activities, depressive and anxiety symptoms, home hazards, foot deformities, mobility assessments, history of falls and consequences, use of medication that are known to be related to falls, and poor eyesight.

3.2. Univariate feature selection

The data was randomly split between training (80%) and test (20%) using stratification. Then, missing values were imputed with the feature's median and the data was scaled with scikit-learn's RobustScaler[1].

The first feature selection step was to filter out features unrelated to our target variable: whether the patient had fallen in the last six months of the intervention's follow up, that is, we want to classify patients as either faller (True class) or non-faller (False class). We employed mutual information[2] to compute how related a feature is to the target variable, as it does not assume equal means, variances, or normality. Since this method can produce different results on each run due to randomness, we averaged the results of 100 runs. Also, this measurement requires a `n_neighbors` parameter, left at the library's default value of 3. We selected features with an average mutual information greater than 0.025, value chosen empirically to filter out approximately half of the features.

3.3. Model selection

After filtering out features unrelated to the target variable, Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) were evaluated using the parameter values in Table 1 to identify the best combination for each model. Evaluation consisted of a repeated (5 times) stratified 5-fold cross-validation. Random forest also ran through 5 random different seeds to account for its variability. We used the *Receiver Operating Characteristic - Area Under the Curve* (ROC-AUC) metric and selected the models based on their average split score and variance, and their consistency across all weeks.

3.4. Recursive feature elimination

The next step was to apply recursive feature elimination using the best parameters for the chosen models. The number of features was chosen using 5-fold stratified cross-validation

[1] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html

[2] https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html

splits, with the ROC-AUC metric. Since some model implementations include stochastic components, we repeated the process 5 times and manually selected a cutoff point for how many times the feature appeared.

4. Results and discussion

4.1. Performance on univariate feature selection and model parametrization

The mutual information filter selected 38, 36, and 44 features for weeks 0, 16, and 22. All of them selected fear of falling (FESI), reported pain, fracture risk (FRAX), poor eyesight, income, health insurance, sex, weight, number of morbidities, foot deformities (Manchester scale), and memory. Weeks 16 and 22 selected fall history in the previous 12 months, but week 0 doesn't have that feature.

The classifiers were then evaluated using the selected features. The selected parameter combinations are underscored in Table 1. Their average cross-validation scores ranged from 0.65 to 0.86, depending on the model and the week in which it was trained. The average cross-validation standard deviation ranged from 0.11 to 0.21, indicating that these results might not be consistent for a dataset of this size and number of features. Random forest was the most consistent model, with deviation ranging from 0.11 to 0.16.

4.2. Recursive feature elimination

The most relevant features were selected after applying RFE with the best parametrization for each model. Table 2 shows the best feature selections, along with the cutoffs for how many times the feature occurred — recall that the models underwent repeated stratified 5-fold cross-validation to reduce variability due to data partitioning. To calculate the "best" score, we considered the average AUC minus its standard deviation across splits.

In Week 0, 34 features were selected by SVM, by far the largest subset. Weeks 16 and 22 selected only 5 and 3 features, respectively. Despite these differences, all models achieved reasonable cross-validation performance, particularly Random Forest in week 22, with both a high ROC-AUC and relatively low deviation across splits (Table 2). Week 0's most important features (calculated by SVM coefficients) were, in order of importance, home hazards, depression/anxiety (part of EuroQol-5D), fear of falling, income, mobility, memory, health insurance, and physical activity (BHPAQ's leisure-time index). The features selected by LR for week 16 were home hazards, anxiety/depression, income, hospitalizations, and pain. The features selected by RF for week 22 were fear of falling, mobility (Timed Up and Go Test), and cognition (ACER).

The results on the separated test set were ROC-AUC 0.71 for week 0 (precision 0.6, recall 0.71), ROC-AUC 0.5 for week 16 (precision 0.4, recall 0.5), and ROC-AUC 0.75 for week 22 (precision 0.57, recall 1.0). Across all evaluated weeks, recall was higher than precision, indicating the models tend toward false-positive predictions. Weeks 16 and 22 presented low test-set performance, which may indicate some degree of overfitting,

Table 2. Cross-validation results from best feature sets

Week	Model	Score	Occurrences	Number of features
0	SVM	0.71 ± 0.15	5	34
16	LR	0.77 ± 0.11	5	5
22	RF	0.89 ± 0.09	3	3

although the relatively high standard deviation could also explain this behavior. This is particularly true for week 16, whose ROC-AUC score is equivalent to random guessing. However, the second best feature set for this week, obtained with SVM, achieved a cross-validation ROC-AUC of 0.72 and a test ROC-AUC of 0.71, with precision of 0.6 and recall of 0.75. It selected home hazards, anxiety/depression, mobility (also part of EuroQol-5D), income, health insurance, and hospitalizations – pretty similar to the other set.

Even though weeks 16 and 22 selected a few features, this may be due to the models' limitations (e.g., the SVM used only a linear kernel) and insufficient samples to learn more complex patterns involving more features. Some of the final features are well documented in literature (e.g., mobility and cognition ones) [Saunders et al. 2025]. At the same time, income, a not-so-studied feature [Saunders et al. 2025], was selected in weeks 0 and 16, and the fall history, which is known to relate to falls [Maruf et al. 2025], wasn't selected in any week. SVM and Random Forest managed to achieve some prediction success while using significantly different feature sets, which could indicate that different stages of intervention require different weights on the assessments, or even, that different case management decisions and slightly different datasets could both lead to different models that look into separate aspects of falls while still achieving reasonable results. The unexpected inclusion and non-inclusion of features, as well as the overfitting behavior of LR, could also be a reflection of limitations stemming from the study's small sample size.

5. Final considerations and future work

While the sample size of 50 patients is not sufficient to confidently generalize the conclusions, we believe this study points toward a promising direction for understanding which factors should be prioritized when using limited intervention time and resources. The proposed approach may help specialists identify assessments that do not contribute enough to falls to justify their cost and effort or assessments and underestimated factors, ultimately leading to more cost-effective and successful interventions.

Overall, the models achieved reasonable AUC scores with Recursive Feature Elimination, although they tended to produce false positives. Feature sets varied substantially across weeks, suggesting that the models may capture different aspects of fall risk over the course of the intervention, or that certain risk factors become more relevant at different stages. The selected features included both well-established and less explored fall-related factors, highlighting potential directions for future investigation. Future work could also investigate prediction models and feature importance in multifactorial and case management interventions using larger cohorts, in order to validate these findings and better understand how fall-related risk factors evolve throughout the intervention process.

6. Acknowledgments

This research was developed as part of an Undergraduate Research Program (*Iniciação Científica*) at ICMC/USP. It was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Finance Code 001), São Paulo Research Foundation (FAPESP, grants 2025/05586-0, 2025/26222-7, 2024/10699-6), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, grant 408323/2024-8), and University of São Paulo (PRPI Ordinance No. 1032, Process 22.1.09345.01.2). Dr Paulo H. S. Pelicioni is currently funded by the National Health and Medical Research Council Investigator Grant (Early Leadership, ID 2042486).

Artificial Intelligence Usage

ChatGPT and Grammarly were used for grammatical revision. Claude assisted with the programming part of the work (syntax, debugging).

References

Alberto, S. N., Ansai, J. H., Janducci, A. L., Florido, J. V. B., Novaes, A. D. C., Caetano, M. J. D., Rossi, P. G., Tavares, L. R. C., Lord, S. R., and Gramani-Say, K. (2022). A Case Management Program at Home to Reduce Fall Risk in Older Adults (the MAGIC Study): Protocol for a Single-Blind Randomized Controlled Trial. *JMIR Res Protoc*, 11(6):e34796.

Capodici, A., Fanconi, C., Curtin, C., Shapiro, A., Noci, F., Giannoni, A., and Hernandez-Boussard, T. (2025). A scoping review of machine learning models to predict risk of falls in elders, without using sensor data. *Diagnostic and Prognostic Research*, 9(1):11.

Ebrahimi, A., Wiil, U. K., Naemi, A., Mansourvar, M., Andersen, K., and Nielsen, A. S. (2022). Identification of clinical factors related to prediction of alcohol use disorder from electronic health records using feature selection methods. *BMC Medical Informatics and Decision Making*, 22(1):304.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422.

Hopewell, S., Adedire, O., Copsey, B. J., Boniface, G. J., Sherrington, C., Clemson, L., Close, J. C., and Lamb, S. E. (2018). Multifactorial and multiple component interventions for preventing falls in older people living in the community. *Cochrane database of systematic reviews*, (7).

Li, Y., Hou, L., Zhao, H., Xie, R., Yi, Y., and Ding, X. (2023). Risk factors for falls among community-dwelling older adults: A systematic review and meta-analysis. *Frontiers in medicine*, 9:1019094.

Liang, H.-W., Ameri, R., Band, S., Chen, H.-S., Ho, S.-Y., Zaidan, B., Chang, K.-C., and Chang, A. (2024). Fall risk classification with posturographic parameters in community-dwelling older adults: a machine learning and explainable artificial intelligence approach. *Journal of neuroengineering and rehabilitation*, 21(1):15.

Maruf, M., Haque, M. M., Hasan, M. M., Farhan, M., and Islam, A. (2025). State-of-the-art review on fall prediction among older adults: Exploring edge devices as a promising approach for the future. *Measurement: Sensors*, 39:101878.

Remeseiro, B. and Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112:103375.

Saunders, S., D'Amore, C., Hao, Q., El-Moneim, N. A., Richardson, J., Kuspinar, A., and Beauchamp, M. (2025). Risk factors for falls in community-dwelling older adults: An umbrella review. *J AM MED DIR ASSOC*, 26(9):105765.

Zhang, H., An, Y., Song, M., and Meng, Y. (2025). Dynamic fall risk prediction in hospitalized cancer patients: development and validation of a machine learning model using multidimensional clinical data to overcome over-sensitivity in traditional scales. *BMC Medical Informatics and Decision Making*, 25(1):377.