Received 14 July 2023, accepted 25 July 2023, date of publication 27 July 2023, date of current version 15 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3299489

RESEARCH ARTICLE

Machine Learning Techniques Applied to the Development of a Fall Risk Index for Older Adults

ANTOINE MILLET[1], ALICIA MADRID[2], JUAN M. ALONSO-WEBER[2], LEOCADIO RODRÍGUEZ-MAÑAS[3], AND RODRIGO PÉREZ-RODRÁ-GUEZ[4]

[1]Biomedical Research Foundation, Getafe University Hospital, 28905 Getafe, Spain
[2]Department of Computer Science, University Carlos III of Madrid, 28911 Leganés, Spain
[3]Geriatrics Service, Getafe University Hospital, 28905 Getafe, Spain
[4]Intelligent Robotics Laboratory, Rey Juan Carlos University, 28942 Fuenlabrada, Spain

Corresponding author: Rodrigo Pérez-Rodrá-Guez (rodrigo.perez@urjc.es)

This work was supported by the Public Spanish Research Funds under Grant PID2019-108408RB-C22-ActiveUP.

ABSTRACT Falls are a leading cause of unintentional trauma-related deaths worldwide, and a significant contributor to elderly dependence. To address this, the goal of this project was to predict recurrent falls in the older population using machine learning techniques, with the aim of reducing the number of falls and their consequences. To achieve this, a dataset obtained from Getafe University Hospital's Geriatric Falls Unit was used (obtained from the Hospital's Electronic Health Records). This extensive dataset was one the key strengths of our work. Feature extraction was performed through natural language processing, which recognized pre-defined patterns and helped build the profiles of the 304 older adults who composed the dataset. The proposed data system was comprised of four main blocks: the senior's profile and environment, clinical information and tests carried out in the hospital, medications, and different diseases they presented. Using the extracted attributes and data from those 304 older adults, this project compared the performance of various machine learning techniques in their ability to classify older adults between future fallers and non fallers. Training different models and ensembles and comparing the results, we obtained that Bagging with Random Forest as base model is the best classifier, predicting accurately 75.8% of the data with 70.0% sensitivity and 80.5% specificity. Ultimately, this research project aimed at setting the first stone to a larger study that could help monitoring older adults and obtaining dynamic and automatic predictions of falls.

INDEX TERMS Fall risk prediction, knowledge-based models, older population, frailty, machine learning, EHR datasets, natural language processing.

I. INTRODUCTION

As of 2021, Falls are being part of the top 3 leading causes of preventable injury-related deaths worldwide. Around 684 000 individuals die from falling each year [1]. Although the mortality rate of a fall-related injury is low, the population having the highest rate of deaths in all regions of the world is people over 60+ years of age. Each year, 28-32% of the senior population over 65 fall at least once [2], and this rate increases to 50% for people over 80. By 2030, 1 in 6 people will be 60 or over, making falls a growing public health concern. In the US only, each year around $50 billion is spent on medical costs related to fatal and non-fatal fall injuries and those numbers will only increase. Healthy individuals are able to adapt more easily to unplanned environmental perturbations such as walking on uneven surfaces or slipping, tripping on hazardous objects. However with aging, age-related pathologies diminish this ability to adapt on-the-go while walking.

With falls come, if non-fatal, along with the cost of hospitalization and medical care, important physical and psychological consequences on the individual that experiences one. According to the Centers for Disease Control and Prevention (CDC), for older adults (aged 65 and older) falling once doubles the chances of falling again [33]. In fact, the risk of falling again within six months is estimated to be as high

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

as 50%. It can result in fractures (often hip fractures) and traumas but it can also affect the senior adult's ability to live alone, partly due to a loss of self-confidence in performing tasks in an autonomous way. This leads to the need of constant monitoring from caregivers or family, which is often not possible for financial or practical reasons. Technology progress has enabled the implementation of assisted-living automated systems (ALSs) [3] in homes to ensure continuous monitoring, providing a safer environment and reducing caretakers workload. However they face different challenges such as acceptability and privacy issues. Regarding the issue of falling in the older adult population, there exist Fall detection systems (FDS), which are part of ALSs, designed to identify falls when they're happening and to inform caretakers so that they can intervene as quickly as possible at the fall's location. Nonetheless FDS has many computational and practical ''hurdles'' such as reliability, real-time processing, sensor placement or even privacy and ethical considerations. Those systems work more as solutions ''in reaction to'' rather than to prevent falls from happening in the future.

Understanding the multifactorial nature of risk factors helps assess evidence-based fall risk of a senior through different techniques that can help take the correct measures to prevent future falling outcomes. In order to classify the senior adult into the fallers (high risk) or non-fallers (low risk) group categories, there exist a variety of prediction models that are based on clinical tests, physical and mental monitoring data and medical records. Falls often involve a variety of interrelated intrinsic and extrinsic factors [25].

Throughout the years, studies have identified many risk factors in the older adult population such as the decline in cognitive function, depression, dizziness, high blood pressure, sleep problems, urinary incontinence, vision impairments, weaker lower-body strength, polypharmacy, DBI drug use (ex: cumulation of sedatives) but most of all balance disorders and impairment gaits, that are the most consistent predictors of falls [26], [27]. We can also include extrinsic factors such as poor lighting, slippery floors for example. Different clinical fall risk assessment methods have been developed over the years, physical test-based or questionnaire-based like the Fall Risk Assessment Tool (FRAT). However the inadequacy of their predictive performance have been known for over a decade, studies have found [22] that the 6 most common fall risks assessment tools Time Up and Go test (TUG), the Berg Balance Scale (BBS), the Functional Reach Test (FRT), Gait Speed test, Performance oriented mobility assessment and Falls history had area under the curve (AUC) values ranging between 0.5-0.7 with sensitivity and specificity varying substantially across studies (ex: TUG, specificity: 28.4%-96.6%, sensitivity: 10%-83.3%). Given those insufficient predictive performances other ways of assessing high fall risks among independently living older people in primary care have been investigated.

With the recent explosion of the number of parameters available in clinical research (through development of ALS, monitoring sensors for example) the traditional data analysis approach used in clinical research (standard univariate, T-test approaches,etc) has been challenged. As an example, for a classification problem with 20 parameters, 1 in 20 associations could be statistically meaningful but not clinically relevant (α level =0.05) [32]. Traditional statistics being very sensitive with small multidimensional datasets, there is an increasing risk of making false conclusions [28].

Machine learning algorithms help alleviate those limitations, using cross-validations to assess their results. The algorithm trains models using a specific part of the data set that was fed to it (the training set) and defines any intrinsic ''rules'' that could exist. Then it tests the validity of those rules by predicting the state of interest (ex: if a senior adult is a faller or a non-faller) on the unseen rest of the data (test set). Through data feeding, process repetition and tracking performance analysis, if conclusive, the models could then be applied to other older adults.

The exponential increase of computed parameters, development of acquisition systems and general trend of joining information from different modalities, is leading to a legitimate interest in using state of the art machine learning algorithms to assess fall risk. Most of the studies done on the subject are based on wearable sensor data providing significant results in classifying fall risk in senior adults [29], [30]. However this type of approach has limitations due to lack of monitoring equipment, caretaker's lack of training, older adult acceptability. A prediction method based on data records only would allow more flexibility and no gear use. Reliability of results is also a limitation of current fall risk assessments, testing different machine learning prediction algorithms and finding an optimal result could help strengthen already proven approaches or provide material for other new research directions.

In this research project, we developed a fall risk prediction algorithm for recurrent falls based solely on the senior adult's profile and environment, clinical information, clinical tests, medicine intake and falls history. As mentioned above, recurring falls are the real problem when looking at falling cases in older adults. The repetition of falling acts increases the risk of falling again, the process of physical deterioration and the fear factor. Our classification algorithm was elaborated in order to predict those recurring falls. The data was extracted for the purpose of this research using text mining on records provided by the Falls Unit of the Geriatric Service of the Getafe University Hospital and relying solely on this data to provide future recurrent falls classification results. The classification system was constructed through the trial of varied ML strategies, defining an optimized method to answer the research question: Can we further optimize future recurrent falls classification in older adults while applying machine learning models solely on profile and environment, clinical information, clinical tests, medicine intake and falls history?

Even though this research work was designed and developed by the technology experts within the research team, clinical experts in the field of recurrent falls among the older population, have been present during the whole process, mainly in the feature selection process as well as in the data interpretation. This interdisciplinary work was carried out during periodic meetings where data scientists presented the results they were obtaining to gather insights to refine/continue their work.

II. RELATED WORK

In order to develop a model that can be effective and used for further research and applications in the future we had to look at the state of the art on the subject. The most relevant methods used in clinical practice and the most recent research done on the subject. Looking at their strengths and weaknesses and trying to address those issues through the system developed in this project.

A. METHODS USED IN CLINICAL PRACTICE

In 1990, Schmid [31] proposed one of the first fall prediction model based on case control using statistical methods. Using data on 204 older adults with very similar ages, separating them in 2 groups: 102 older adults that had suffered falls in the last month and 102 older adults that had not suffered falls.

In 1995, another relevant study was undertaken by Hendrich et al. [19] of which the prediction model is still used by nurses in clinical settings today. A fast and effective test (77%sensitivity), based on statistical methods that was developed with the help of 102 older adults whose falls had occurred in the last month and 236 random people in the control group, providing greater variability of individuals than Schmid.

Subsequently, in 1997, the most used method for predicting the probability of falling in clinical settings was released: the Morse Fall Scale (MFS) [20]. This test became a reference for its efficiency in application (maximum time of three minutes), straightforwardness (score of different variables) and very favorable results (83% specificity). Morse followed a similar strategy to Hendrich II, using a statistical methodology, balancing both groups using 100 individuals with past falls and 100 random people with no past falls. The predicted risk referred to a prediction of approximately four months

That same year, Oliver et al. [21] developed another fall model using variable analysis (statistical method) known as the St.Thomas risk assessment, which is widely used nowadays in clinical practice (91.5% sensitivity). Oliver applied some population filters on his selected population, limiting the sampling to individuals older than sixty-five (116 fallers and 116 non-fallers). The test provided a prediction time of approximately three months.

All tests mentioned had a point based scaling based on different variables. Differences existed in the variables chosen to develop their test but they all agreed on common risk factors that are mobility/gait and mental state. Hendrich et al.'s work [19], building on Schmid's research while using five risk factors (mobility, mental state, bathroom use, previous falls and medication), also mentioned other risk factors such as dizziness or gender.

TABLE 1. Results of clinical methods.

Method	Sensitivity	Specificity
Schmid et al.	95%	66%
Henry II et al.	77%	72%
Morse Fall scale	78%	83%
Oliver et al.	91%	78%

B. SIMILAR RESEARCH

In addition to the methods used in clinical practice, it is also necessary to analyze current research approaches that involve state of the art machine learning methods in their predictive analysis. As it was mentioned above, with the recent burst of medical data available, traditional data analysis used in clinical research has been challenged and researchers are now looking for up to date methods that can make the most out of the extensive data available. Looking through the literature, we found out what type of information, data features and machine learning techniques were being used, the objectives set and the results obtained.

Hosseini et al. [13] conducted a study to predict the risk of repeated falls in older adults using a predictive model that combined vital signs, medical history, and movement data obtained through accelerometers. The study compared the model's results with the MFS scale, but the sample size was too small (30 people over 50 years old) to confidently validate its effectiveness, despite achieving good results in sensitivity and precision.

Zhang et al. [7] applied five learning algorithms to a study with a population of only 35 people over 60 years old, making it, similarly to the Hosseini study, difficult to generalize the results. The algorithms used were K-Neural Network, Naïve Bayes, Logistic Regression, and Support Vector Machine. The study also included a 10-fold cross-validation that predicted the first fall within 12 months.

Trying to go past the small population sample issue, Greene et al. [8] conducted a study over 5 years (2007 to 2012) to obtain enough data (616 older adults over 70 years old), combining cognitive impairment data with informations extracted from sensors, generating a more generalized model using logistic regression. The model obtained a sensitivity of 70% and specificity of 39%, which implied that the model mainly predicted that a given older adult would fall.

The previous studies mentioned all combined sensors reading with clinical information. However, even though they provide a broader data collection, sensors require a higher budget to develop accurate models. In the work of Noh et al. [5], similar sample quantity to Greene et al. [8] was used (620 entries with 210 fallers and 410 non fallers between the age of 55 and 97 years old). The variables of the study only included older adult's clinical information (disease and medications) but

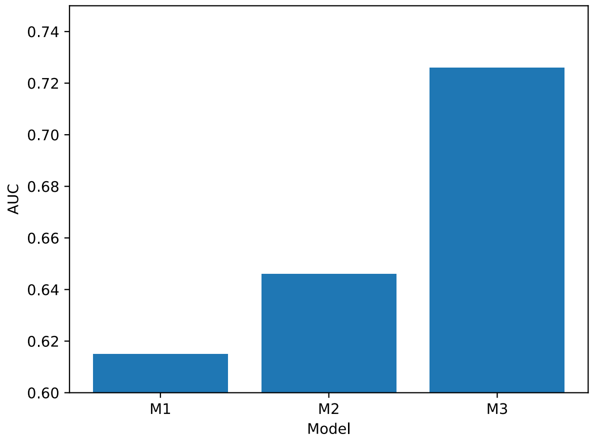


FIGURE 1. AUC comparison of models M1, M2, M3.

their analysis was much deeper, dividing attributes into different groups: MFS scale attributes (M1), similar covariance attributes (M2) and older adult's conditions (M3) that were progressively added. Looking at the results, we observed that a higher number of variables led to increased results.

Previous fall risk assessments mentioned used few variables (5-12) in each test. However some studies [2] carried-out a much more in depth analysis, looking at polypharmacy, time-related fall risk. The Agency for Healthcare Research and Quality [5] also mentioned education levels or disease evolution status as potential determining features. Our study aimed at analyzing a wider range of risk factors using multiple domains together, including personal and clinical information, medicines, and diseases. One of the major strengths of our study was that we had access to very precise health records information from people who had experienced falls and were part of the Falls specialized unit of the Geriatrics Service of the Getafe University Hospital. This has led to a large amount of variables separated in 4 categories: personal, clinical information and test results, medical conditions and medications taken. In order to select features that had already been mentioned as relevant in the past, confirmed by experts in the field and by research results, we looked at past papers on the subject, analyzing which features had been selected by others. The selection of features is displayed in Table 2.

Our research work used a machine learning approach, automating the entire process and being a novel approach compared to current clinical practice. In order to define an AI algorithmic approach that could tackle a large amount of varied features, we had to look at similar studies with large set of variables like ours. Specifically the work of Chu et al. [32], Lindberg et al. [11], Jung et al. [10], Liu et al. [12] and Chan et al. [6]. Their machine learning approaches and outcomes are summarised in Table 3 below.

III. MATERIALS AND METHODS

A. METHODOLOGY FLOWCHART

The methodology that we followed in this study is detailed in the Figure 2 below.

TABLE 2. Feature selection of similar research papers.

Feature	Hosseini et al.	Zhang et al.	Greene et al.	Noh et al.	Chan et al.
Personal information					
Age	X		X		
Gender	X		X		
Height			X		
Weight			X		
Drugs	X				
Alcohol	X				
Clinical information					
Disorientation	X				X
Cognitive deterioration	X		X	X	X
History of falls	X				X
Grip force	X	X			
Gait speed	X	X			X
Stride length	X	X			
Balance	X	X			X
Sole maximum inclination		X			
Irritability					X
Dizziness					X
Walking habits					X
Medical conditions					
Hypotension	X		X		
Diabetes				X	
Dementia				X	
Hypoalbuminemia				X	
Hyponatremia				X	
Medicines					
Polypharmacy	X		X		
Calcium channel blockers				X	X

TABLE 3. Results and machine learning approaches of similar papers.

Reference	Population	Number of features	Machine Learning model	Results (AUC)
Chu et al.	1115	15	Boosting+Decision Tree	0.72
Lindberg et al.	814	38	Bagging+Random Forest	0.91
Oshiro et al.	57 678	13	Logistic Regression	0.74
Jung et al.	5129	165	Decision Tree	0.73
Liu et al.	15	54	Bagging+Random Forest	0.74
Chan et al.	602	10	Stacking+Decision Tree	0.70

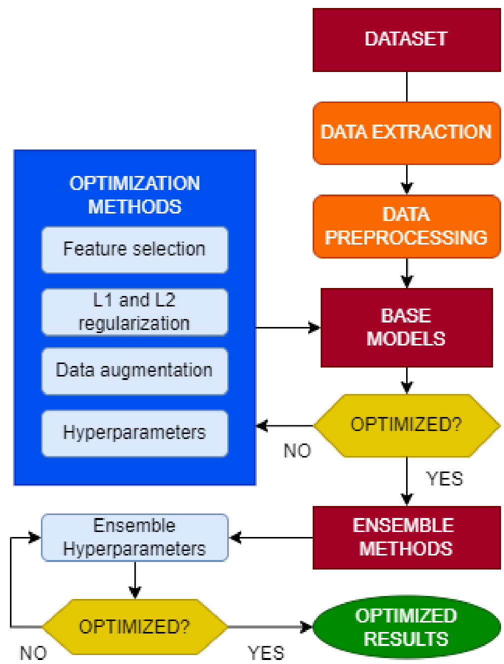


FIGURE 2. Flowchart of methodology.

B. DATASET

Data were collected in clinical routine, following the data collection protocol defined in the Hospital's Geriatric Falls Unit. These data were not collected specifically for the experiments herein described but earlier in time, with the initial intention to be only used for clinical purposes, including diagnosis, prognosis and intervention design. To collect these data, the software supporting Getafe University Hospital's Electronic Health Record system was used. This application consists on different forms and text boxes where both results obtained for each tests and practitioners' comments can be added. It is important to highlight that, even though there are guidelines to collect data for all assessment tests, practitioners do not always use them, however they often include information in the text boxes. This justified the effort the research team had to put into the natural language processing aspects to extract valuable information.

Our research was based on the data set provided by Getafe University Hospital's Geriatric Falls Unit (obtained from the Hospital's Electronic Health Records). When an older adult experienced a fall that required medical attention they went to this unit in order to initiate personalized treatment and monitoring to prevent repeated falls. From the point of admission, older adult's data was collected during clinical routine, following the data collection protocol defined in the Hospital's Geriatric Falls Unit.

These data were not collected specifically for the experiments herein described but earlier in time, with the initial intention to be only used for clinical purposes, including diagnosis, prognosis and intervention design. At no time there was a guideline produced specifically for the project, as we tried to develop an instrument using the usual data collected by the clinicians in their daily and routine practice.

Data were updated in every follow-up visits regarding the fall risk factors which had been managed in the intervention: new drugs or doses, functional markers if a physical exercise program had been implemented, conditions of the environment if some changes had been promoted….In the case of a second and successive falls, all the data were re-assessed including, among others, the characteristics of the fall, the setting, the environment, the injuries produced, the potential new factors and the changes in other ones previously recorded (drugs regime, cognitive or functional deterioration), etc. The data collected for this research was mainly written in natural language, maintaining a semi-structured design which is quite common in medical records as is discussed in Hossain and Soar's work [34]. In addition, a series of clinical tests were included in a numerical format.

It is important to highlight that, even though there are forms to collect data for all assessment tests, practitioners do not always use them. However, they include much information in the text boxes. This justified the effort that our research team had to put into the natural language processing aspects to extract valuable information.

Pseudonymization was implemented to protect each older adult's anonymity, replacing their first and last name with a unique ID. Data referring to the individuals such as name, address, etc. were initially removed from the dataset so entries were only identified by their Electronic Health Record identifier. Moreover, these identifiers were matched with other random identification numbers to decouple the information. The principal investigator kept the association table between official IDs and random ones in case there was a need of going backwards (this method was approved by the ethical committee granting this study). This study was approved by the Drug Research Ethics Committee of the Getafe University Hospital with the following code: CEIm2137.

As mentioned, the vast majority of the data was provided in natural language, with semi-structured data and little specific and practical information for an automatic learning model. Thus it has been necessary to perform an extraction of attributes and/or characteristics from the initial data provided.

In our initial selection process, we investigated the scientific literature [5], [6], [7] [8], [13], some features being mentioned in Table 2, we analyzed the works of Lord et al. [35], Rubenstein [36] which discussed the importance and relevancy of a majority of the features we ended up selecting. We were also able to obtain precious insights from geriatricians that worked in the Falls Unit of the Geriatric Service of the Getafe University Hospital and that had been specially trained to treat recurrent fallers, thus having relevant knowledge on the features we should include in our study.

All of those factors combined helped us determine which attributes were relevant to our research. In this way, the set of attributes was divided into four main groups: older adult profile and environment, clinical information and tests carried out in the hospital, medications taken and the medical conditions they were intended to cure or alleviate.

C. DATA EXTRACTION

Natural language processing was implemented on the data records, in order to detect keywords or patterns and extract the desired attributes. The tool used through the process was Spacy [37]. In addition to that, in the case of medications and conditions, text mining has been used to obtain the most recurring words. Tools used have been Scikit-learn [38] with the TfidfVectorizer class that transforms a set of documents into a TF-IDF matrix. Overall process of text mining for feature extraction is explained in the flowchart of Figure 3

In each iteration of the loop, that is, in each selected row (entry in the dataset), a dictionary was initialized with default values (0 in numeric attributes and "Null" in boolean or nominal values) which contained all the variables related to the given data group (for example: profile and environment of the older adult). A first loop is done, trying to retrieve as many variables as possible. Subsequently, if the dictionary is not complete, each of the columns of the row in question were analyzed again looking for the attributes whose value had not yet been found. Finally, once all the values or columns were checked and completed, the data was saved in another dataframe along with the older adult's ID and the date of entry.

All the attributes were extracted through a function programmed in Python, in which inconsistencies are corrected (spaces, transformation to lower case, elimination of tildes, etc.). In each run, the patterns that existed in the document

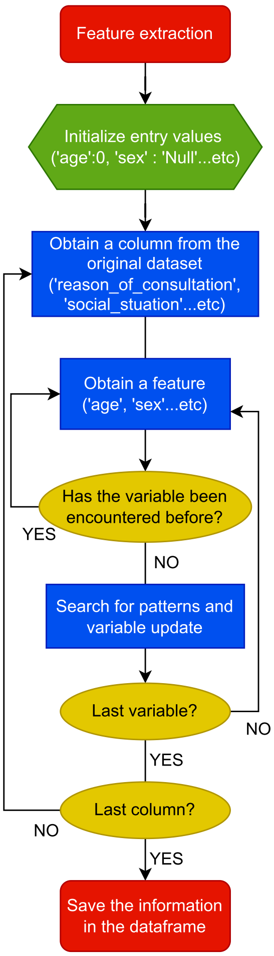


FIGURE 3. Flowchart for feature extraction.

were identified and analysed one by one in reverse order, that is, always starting with the last pattern. Then, each of the tokens that made up the patterns were checked in order to extract the specific value of the attribute.

To determine the class value for the predictive model, it was analyzed whether there were falls that occurred after the older adult's first visit. This value had to be a boolean. If there were any falls that had taken place, the value of the class was set to True, and it was possible to obtain the approximate time that elapsed between one fall and the next. On the other hand, if there were no falls recorded after the senior adult's first visit, the class value was set to False. This information was crucial for the prediction of falls and the implementation of preventive measures that can significantly improve senior adult safety and quality of life.

D. DATA PREPROCESSING

The data extraction has resulted in 96 features, which included the senior adults' profile and environment, their clinical information and results of clinical tests carried out, medications taken and conditions. Pre-processing of these data was necessary before constructing the learning system. Effective data cleaning was essential for developing accurate and reliable machine learning classifiers, as errors or inconsistencies in the data can significantly impact the performance of the model.

1) DATA CLEANING

First, data's plausibility was verified so that any type of data that contained noise or incorrect values be eliminated. For example, for the age feature, all values under 50 and over 110 were deleted, since the clinic did not see senior adults outside this age range.

2) DATA TRANSFORMATION

Normalization of the numerical data is done in such a way that the range of values is always between 0 and 1 for all features. The implemented formula is the following:

$$X_{\text{norm}} = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{1}$$

where X is the original variable, and X_{norm} is the normalized variable.

This technique was applied since, for example, the range of values for age was [50,110], while for BMI it was [0,2]. In this way, the range of values were equal which facilitated the learning process. Finally, to finish the transformation of the data frame, the boolean values (True and False) were changed to 1 and -1 respectively, since Scikit-Learn only accepts numeric values in its algorithms.

3) DATA REDUCTION

Data reduction refers to the process of reducing the amount of data used in a machine learning model, while still retaining the most relevant and informative parts of the data. There are several reasons why data reduction may be desirable, such as reducing the computational requirements of the model, improving the accuracy of the model, and reducing the risk of overfitting.

The original data set obtained from the Falls Unit of the Geriatric Service of the Getafe University Hospital was requested on December 15, 2021, so the data had been collected until December 13 of that same year. The extraction of falls explained in the previous section, included falls that occured between January 2017 and December 2021. However, if a senior adult fell in December 2021, it was impossible to know if said older adult had suffered or will suffer a new fall, since not enough time had elapsed for the individual to come back for a follow up or exceptional post-fall visit and update his profile. Therefore, the data was reduced, so that all the entries whose fall was after September 2021 were eliminated, so that there was a reasonable time from the fall to data extraction.

E. FEATURE SELECTION

Furthermore, we used feature selection, which involved selecting a subset of the original features (also known as variables or predictors) that were the most relevant to our fallers

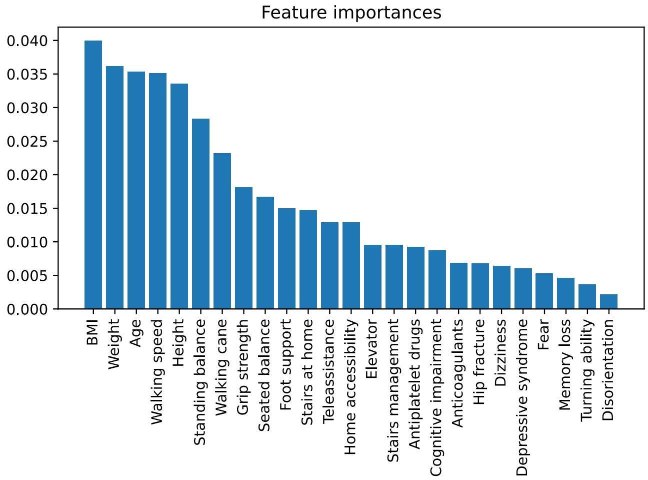


FIGURE 4. **Top 25 feature's importance.**

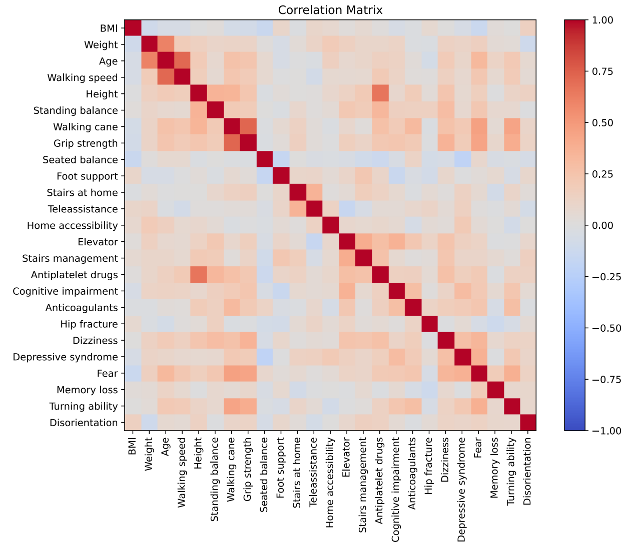


FIGURE 5. **Top 25 feature's correlation matrix.**

classification problem. Feature selection can be particularly important when dealing with high-dimensional data, where the number of features is very large relative to the number of data points. This is explained in section II-A of Guyon and Elisseeff's work [39], where high-dimensional data is demonstrated as leading to overfitting and increased computational complexity, it develops on how feature selection can help address these issues. In our case, geriatricians had agreed on an initial set of features they considered the most relevant given their experience, which were actually most of those collected in routine care since their data collection process were specifically designed for falls assessment. The current feature selection carried out corroborated and refined their selection, always considering their feedback, obtaining 96 features and a complex overview of each senior adult involved in the research. This feature selection was done to observe what features out of those 96 original features collected were the most decisive. In order to study the influence of each feature on the fall risk assessment as well as trying to optimize our future results, we looked at calculating feature importance using Random Forest as it had been proven effective in similar papers like in the work of Acharjee et al. [43] where Random Forest is used to discover biomarkers for early diagnosis of diseases or the work of Yadav and Pal [42] that uses RF to select features in order to predict heart diseases, among others [40], [41]. Feature importance was calculated by measuring how much the performance of the model decreased when a particular feature was randomly permuted. The basic idea was that if a feature is important for making accurate predictions, then randomly permuting its values should cause a significant decrease in the performance of the model. Results obtained using this method are displayed in Figure 4.

We can observe on this table that Body Mass Index (BMI), Weight and Age were the main deciding factors in a fall risk assessment, those variables could be considered as ''umbrella'' terms as they encompass other conditions that could potentially influence fall risk index such as mobility, cardiovascular diseases, mental state. Those variables are part of a general consensus [1] such that age and mental and physical states are the most critical factors in a person ability to avoid falls. We can observe that there was a dramatic difference in coefficient value between the most important feature and the feature n°25.

Calculating the correlation matrix of the top 25 features also helped to identify highly correlated features and avoid redundancy. In our case, no clusters of features were experiencing sufficient correlation to affect our models predictions, as can be seen in Figure 5 so no further feature selection was done on the 25 features.

The Results section will explore how reducing the number of features used to train a model can influence the results.

More information on the decisive features that were obtained using Random Forest feature selection are displayed in Table 4 below.

Other ways to favor specific features in order to reduce the complexity of the model were L1 and L2 regularization.

F. L1 AND L2 REGULARIZATION

Regularization in machine learning prevents overfitting by adding penalty terms to the cost function. L1 and L2 regularization are common types of regularization used in classification models [44].

L1 regularization (Lasso regularization) selects important features by adding their absolute coefficients to the cost function [45]. This reduces model complexity and performs feature selection.

L2 regularization (Ridge regularization) encourages small non-zero coefficients for all features by adding their squares to the cost function [45]. It mitigates overfitting by reducing coefficient magnitudes.

L1 and L2 regularization were considered to be implemented as a data-preprocessing in our research as they have been used in past similar research papers [8], [12] and have

TABLE 4. Descriptive data of the 25 selected features.

Older adult's characteristics	Total (n=304)	Fallers (n=129)	Non-Fallers (n=175)
BMI	33.7 (±9.2)	33.4 (±9.5)	34.0 (±9.8)
Weight	71.0 (±19.0)	71.6 (±23.0)	70.2 (±13.8)
Age	80.3 (±7.7)	80.4 (±6.7)	80.2 (±8.4)
Gait speed	0.8 (±0.5)	0.8 (±0.2)	0.8 (±0.7)
Height	169 (±11.2)	170 (±10.4)	168 (±13.7)
Standing balance	123.5 (±48.9)	135.1 (±42.2)	105.8 (±53.4)
Walking cane	183 (60.2%)	67 (51.9%)	116 (66.3%)
Grip strength	16.2 (±6.0)	15.0 (±4.0)	17.4 (±7.2)
Sitting balance	137.8 (±26.9)	139.9 (±24.7)	134.7 (±29.9)
Foot support	133 (43.8%)	63 (48.9%)	70 (40.0%)
Stairs (home)	94 (30.9%)	41 (31.8%)	53 (30.3%)
Tele-assisstance	141 (46.4%)	54 (44.2%)	87 (49.7%)
Home accessibility	208 (68.4%)	81 (62.8%)	127 (72.6%)
Lift	53 (17.4%)	23 (17.8%)	30 (17.1%)
Stairs management	94 (30.9%)	41 (31.8%)	53 (30.3%)
Antiplatelets drugs	81 (26.6%)	46 (35.7%)	35 (20.0%)
Cognitive impairment	69.0 (22.7%)	32 (24.8%)	37.0 (21.1%)
Anticoagulants	50 (16.4%)	17 (13.2%)	33 (18.9%)
Hip fracture	137 (45.1%)	55 (42.6%)	82 (46.9%)
Dizziness	61 (20.1%)	28 (21.7%)	33 (18.9%)
Depressive syndrome	83 (27.3%)	41 (31.8%)	42 (24.0%)
Fear	56 (18.4%)	35 (27.1%)	21 (12.0%)
Memory loss	43 (14.1%)	19 (14.7%)	24 (13.7%)
Stable turns	21 (6.9%)	11 (8.5%)	10 (5.7%)
Disorientation	47 (15.5%)	27 (20.9%)	20 (11.4%)

proven to be effective. The Results section will develop on how they influenced our results.

G. DATA AUGMENTATION

Another way to boost the performance of the models was to apply a data augmentation to our training set. This method proved to be effective in similar past papers [46], [47]. Data augmentation is a technique used to increase the amount of training data available for a machine learning model by creating new, modified versions of the existing data. In the context of our classification model for classifying high and low fall risk among senior adults, data augmentation involved creating new training examples by applying various transformations to the existing training dataset. There exist different types of data augmentation approaches, such as noise injection or feature mixing, used in the work of Martins et al. [48]. In our research work, we used feature interpolation. Generating new examples by interpolating features between existing records, while still preserving the normalized scale of the data.

H. MACHINE LEARNING AND PREDICTION MODEL DEVELOPMENT

In order to predict the risk of falling in senior adults, we used 6 different base models. The main criterias of selection were based on ability to handle highly dimensional noisy data containing a mix of continuous and categorical values. Use of the model with good results in previous research papers was also taken into account. The classification algorithms were Decision Tree, Random Forest, Logistic Regression, LightGBM, Support Vector Machine (SVM) and K-Neural Networks (KNN). We disregarded models such as multi-layer perceptron or basic linear models, as they had been tested in the first iteration of our research and provided average results [72].

1) DECISION TREE

Decision tree is a machine learning algorithm used for classification and regression tasks. They model decisions or actions by splitting the data into branches based on a set of conditions or rules. New samples are classified by traversing the decision tree based on the values of the features for that sample [49]. We chose to include this algorithm because Decision trees are a popular choice for classification problems as they are easy to interpret and can handle both numerical and categorical data. They are especially useful for data with many features, like medical records. Use of this algorithmic strategy can be found in other similar paper such as in the work of Vasudha Rani et al. [55] and Makino et al. [56].

2) RANDOM FOREST

Random Forest is an ensemble learning method that uses multiple decision trees to improve predictive performance. It reduces overfitting and improves generalization by aggregating predictions from multiple decision trees. Each decision tree casts a vote for the classification result, and the final prediction is determined by selecting the class label with the highest number of votes [50]. We chose to include Random Forest as it is a powerful and widely used machine learning algorithm, particularly for high-dimensional data with complex decision boundaries. It can handle both categorical and continuous features, and is robust to noisy data and missing values. Use of this algorithmic strategy can be found in other similar paper such as in the work of Lindberg et al. [11] and Lyu et al. [57].

3) LOGISTIC REGRESSION

Logistic regression is a linear model for binary classification. It finds the best-fit line to separate the two classes using available data. It predicts the probability of an instance belonging to a particular class and outputs a value between 0 and 1. A threshold value maps the output probability to a binary class label [51]. We chose to include this algorithmic approach as it is simple to interpret and can be used to estimate the probability of an event occurring. It is often used in medical research for risk prediction. Use of this algorithmic strategy can be found in other similar paper such as in the work of Bu et al. [58] and Jung et al. [10].

4) LightGBM

LightGBM is a gradient boosting framework that uses tree-based algorithms to build predictive models. It offers unique features like sparse optimization, parallel training, various loss functions and early stopping. Unlike most implementations, LightGBM grows trees leaf by leaf, selecting the leaf that yields the greatest reduction in loss. It also employs a histogram-based decision tree learning algorithm that provides good performance and memory savings compared to sorted-based decision trees in other implementations. The algorithm further improves model performance with Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) [52]. We chose to include the LightGBM for its efficiency, flexibility and robustness to noise and missing values. Use of this algorithmic strategy can be found in other similar paper such as in the work of Chu et al. [32] and You et al. [60].

5) KNN

KNN stands for K-Nearest Neighbors, a simple yet effective machine learning algorithm for classification and regression problems. KNN is a non-parametric method that works by identifying the K closest training examples in the feature space to a given input, and then classifying or predicting the input based on the majority class or average of the K neighbors [53]. We chose to include this algorithmic approach because it is simple and easy-to-understand, does not require any training process. It also has the advantage of being able to handle multi-class classification problems and can work with both numerical and categorical data. Use of this algorithmic strategy can be found in other similar paper such as in the work of Chu et al. [32] and Gunale and Mukherji [61].

6) SVM

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks. It finds a decision boundary that separates the data into classes, with the objective of maximizing the margin between the boundary and the closest data points [54]. We chose to include SVM as it can handle high-dimensional data with few training examples, and can also handle noisy data and outliers. It allows for control of the trade-off between model complexity and generalization performance, and can also be used for regression analysis. Use of this algorithmic strategy can be found in other similar paper such as in the work of Chu et al. [32] and Yokota et al. [62].

I. CROSS-VALIDATION

In order to evaluate the accuracy of our machine learning models for fall risk in older adult people, we employed cross-validation in order to evaluate the reliability of our models [63]. The dataset was composed of 304 entries. Cross-validation is a statistical method that aims to avoid errors caused by a model's excessive dependence on specific training set. To achieve this, the parent dataset was cut into a greater number of subsets, allowing for different combinations of data sets.

During each round of training, different subsets were selected as the training set and test set, in order to minimize modeling errors. The stratified K-fold cross-validation method was used, which involved dividing the entire data set into K equally sized parts, where K is a user-defined number [63]. In our case, K was set to 10 (so here we have 6 sets of 30 entries and 4 sets of 31 entries). During each of the 10 rounds, one of the K subsets was used as the test set for model verification, while the remaining K-1 subsets were used for training. This process was repeated for each of the K subsets to ensure that each subset was used for both training and testing.

After 10 rounds of training, the accuracy results of the 10 modeling exercises were averaged to provide a fair estimate of the model's performance on the overall dataset. By using cross-validation, we were able to obtain a more reliable estimate of the model's accuracy, as it was not dependent on a single training set.

IV. RESULTS

In order to obtain the best single classification model possible, we decided to train models using the different algorithmic approaches defined in Section III.E (Decision Tree, Random Forest, LightGBM, Logistic Regression, KNN and SVM); approaches that have already been tested in previous studies and provided satisfactory results.

Firstly each of the models were trained over several iterations in order to tune the hyperparameters. The Decision Tree model was left with the default values. The LightGBM classifier was tuned to a learning rate of 0.01. This learning rate is small enough to allow the model to converge slowly and avoid overshooting the optimum solution, but large enough to prevent the model from getting stuck in local optima. The Random Forest classifier was tuned to have 100 estimators as it provides a good balance between accuracy and efficiency for many datasets and problems. The class weight was also changed to a 20 to 1 ratio as we observed that sensitivity was quite low and it needed to be corrected. The SVM classifier was tuned to have a RBF (radial basis function) kernel. The kernel function is used to transform the original feature space into a higher-dimensional space, where it may be easier to separate the classes using a linear boundary or hyperplane. In practice, the RBF kernel is often a good choice for SVMs when the relationship between the input features and the output variable is complex and nonlinear. The LR (Logistic Regression) classifier was tuned to have LBFGS (Limited-memory Broyden-Fletcher-Goldfarb-Shanno) solver as it is generally a good choice for LR when dealing with large datasets or a high number of input features. The solver is an optimization algorithm used to find the coefficients that best fit the model to the training data. The KNN classifier was tuned to have distance-based weights as it can be beneficial when the dataset has non-uniform distribution or class imbalance which was our case (129 fallers for 175 non-fallers).

As mentioned in the Materials and Methods section, we also implemented L1 and L2 regularization, in order to prioritize given features over others and reduce the complexity of the models. We observed an important increase in results as can be observed in Figure 6

Within the logic of simplifying the models and trying to optimize the results, we modified the set of features,

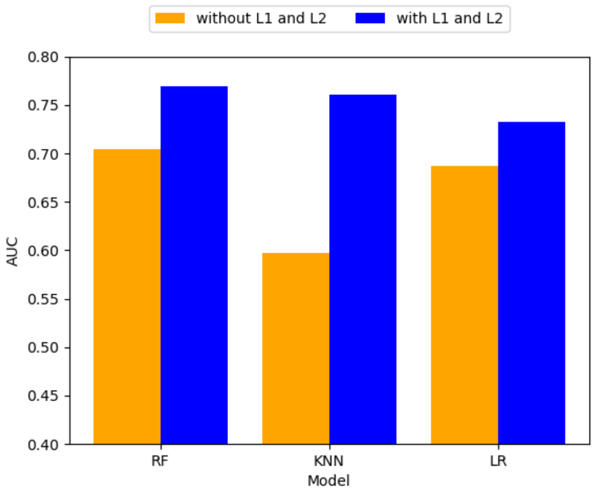


FIGURE 6. Evolution of AUC with and without L1 and L2 regularization.

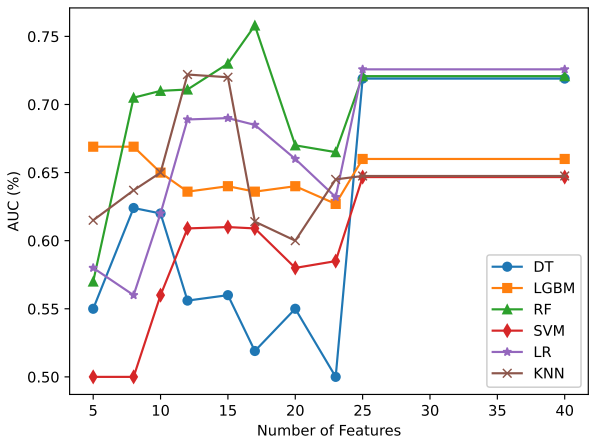


FIGURE 7. Evolution of AUC with number of features.

TABLE 5. Results of base models.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	f1 score (%)	Precision (%)
Decision Tree	71.1	73.3	70.5	71.9	66.2	63.3
LightGBM	69.8	40.5	87.0	66.0	51.8	70.2
Random Forest	**76.9**	**66.7**	**85.0**	**75.8**	**70.1**	**82.5**
SVM	67.1	50.8	78.5	64.7	56.4	73.2
Logistic Regression	73.3	66.7	78.5	72.6	65.4	75.8
KNN	76.1	51.2	91.5	72.8	60.9	71.7

TABLE 6. Standard deviation and result of best base model.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	f1-score (%)	Precision (%)
Random Forest	**76.9**	**66.7**	**85.0**	**75.8**	**70.1**	**82.5**
Standard Deviation (+/-)	4.3	6.1	4.9	6.1	4.7	5.6

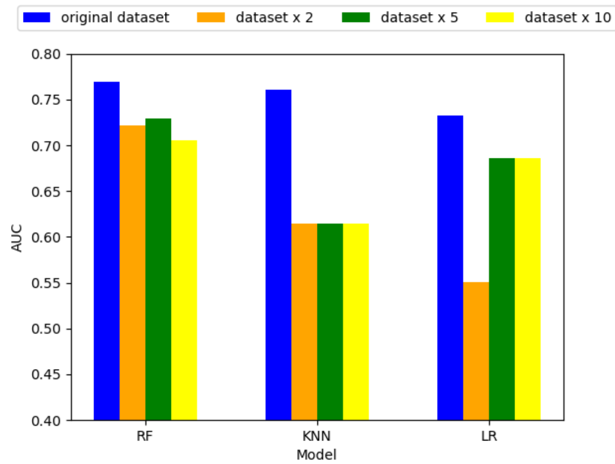


FIGURE 8. Data augmentation AUC results for top 3 models.

removing the least important factors, using Random Forest as a feature selection function. The results obtained for each algorithm approach, on features set of 20, 40, 60, 80 and 96 variables, and displayed in Figure 7 showed a decrease in performance with reduced number of features. This coincided with the conclusions obtained from Noh et al. [5] which had deduced that the higher the number of features the better the results. However it is to be noted that the performances of the models peaked and stabilized past the point of 25 features (except for Random Forest which reached a peak at 17 features). Thus we can deduce that the information that was bringing the rest of the feature set was not relevant or decisive enough to influence in a significant way the results. We decided to only use the top 25 features to reduce complexity and computational process time.

The full results of the classifier models can be found in the table below.

Looking at the cross-validated results obtained after tuning we can see that Random Forest obtained the best overall results with 76.9% accuracy and 85.0% specificity, followed by KNN and Logistic Regression (76.1% and 73.3% accuracy respectively). Those results are understandable as the RF algorithmic approach copes well with high dimensional problems and is robust to noisy data. It has been providing great results in past research too [11], [12]. The standard deviations provided by the averaged metric results are also low which indicates consistency across different folds and suggests a more reliable and stable estimation of the model's performance.

Concerning the data augmentation method that was mentioned in the Materials and Methods section, the results before and after data augmentation were compared and are displayed in Table 8.

We observed that the results decreased while increasing the number of data in the training set. Indeed data augmentation can also lead to overfitting, preventing the model from generalizing the learning to unknown new datasets.

On the process of optimizing the models results, it was necessary to look at ensemble methods. Ensemble methods are machine learning techniques that combine multiple individual models to improve the accuracy and robustness of the overall model [64].

The basic idea behind ensemble methods is that by combining the predictions of multiple models, the strengths of each individual model can compensate for the weaknesses of others, leading to a more accurate and robust overall model [65]. In our case, using ensemble methods had the potential to be useful because we had a relatively small number of senior adult entries compared to the number of features, which could have led to the overfitting of individual models. By combining multiple models trained on different subsets of the data, we were able to reduce the risk of overfitting and improve the generalization performance of our classification model. Additionally, ensemble methods can be robust to irrelevant variables in the data, which is important when dealing with medical health records that may contain many potentially irrelevant features.

We were able to apply 3 ensemble methods, Stacking, Bagging and Boosting, that were tested on each base model. Stacking, Bagging and Boosting are all ensemble methods that combine weak learners in order to improve the overall prediction [64].

In Bagging, each model is trained independently and makes a prediction based on the input features. The final prediction is made by averaging the predictions of all the individual models. Use of Bagging in similar context research can be found in the work of Liu et al. [12] and Lindberg et al. [11]

In Stacking, the predictions of each model are used as input features to a second-level model, which is trained to make the final prediction. Use of Stacking in similar context research can be found in the work of Hwangbo et al. [66] and Nguyen et al. [67]

In Boosting, each model is trained to focus on the samples that were misclassified by the previous models, and the final prediction is made by combining the predictions of all the individual models. Use of Boosting in similar context research can be found in the work of Mahariba et al. [68] and Lindberg et al. [11]

For the stacking ensemble approach we explored combining complementary models (Random Forest and Logistic Regression or SVM and LightGBM classifiers) but also by training the ensemble with only one type of base model. The meta-model for the 2nd level training was Logistic Regression as it is commonly used in binary classification problems [6], [60]. It is easily interpretable and computationally efficient and can boost the results if the base models are diverse and produce accurate and complementary predictions. The number of base models in each stacking ensemble were tuned in order to combine best results and diminish model's complexity. The top 5 model results obtained out of all the ensemble models trained are displayed in Table 7. Those results were cross-validated using a 10-fold cross validation method (here 6 sets of 30 entries and 4 sets of 31 entries).

The best results were obtained using the Bagging ensemble method combined with Random Forest as base model (75.8% accuracy, 70% sensitivity, 75% AUC) which was expected based on the argument above as well as the previous research of Lindberg et al. [11] and Liu et al. [12] who obtained their best performances with the same combination as can be seen in Table 3. The standard deviations provided by the averaged metric results are also low which indicates consistency across different folds and suggests a more reliable and stable estimation of the model's performance as can be seen in Table 8.

This was followed by a Stacking ensemble with Random Forest as base model (76.8% accuracy, 74.5% AUC) and a Stacking ensemble of 4 Logistic Regression models (74.5% accuracy, 73.6% AUC)

TABLE 7. Results of ensemble models.

Ensemble model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	f1-score (%)	Precision (%)
Stacking+RF-LR	73.2	54.2	87.5	70.8	61.1	82.5
Stacking+RF	76.8	60.0	89.0	74.5	67.2	90.0
Bagging+RF	**75.8**	**70.0**	**80.5**	**75.3**	**69.5**	**77.3**
Bagging+LR	74.5	66.7	80.5	73.6	67.7	76.3
Boosting+LR	72.1	66.7	76.5	71.6	63.6	73.5

TABLE 8. Standard deviation and results of top ensemble model.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	f1-score (%)	Precision (%)
Bagging + RF	**75.8**	**70.0**	**80.5**	**75.3**	**69.5**	**77.3**
Standard Deviation (+/-)	5.2	6.0	4.7	5.8	3.5	5.1

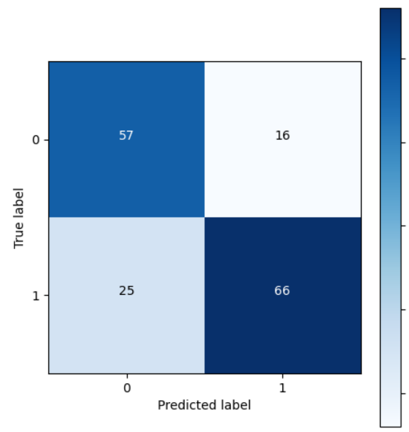


FIGURE 9. Confusion matrix of Bagging with Random Forest model.

V. DISCUSSION AND CONCLUSION

In this study we combined biological, behavioral, environmental and socio economic data of 304 older adults (129 fallers and 175 non-fallers) which resulted in 96 features that were used to train our machine learning models and ensembles and enable them to predict older adults' risk of suffering a recurrent fall. The good performances of those models could be used in future clinical applications to make hierarchical decisions and implement fall prevention strategies to reduce subsequent falls and hospitalization in the older population.

Our study distinguished itself by the high level of data precision that we were able to obtain on each senior adult, implementing natural language processing on medical records coming from the specialized Falls Unit of the Geriatric Service of the Getafe University Hospital. That data extraction provided us with a complex and diverse set of data that we were able to use to train our models but also to study their respective importance in the classifying process. Results from our findings provided that BMI, Weight, Age, Walking Speed and Height were the top 5 features that were determining if a senior adult was going to suffer a fall in the next few months or not. This went in the same direction as the general consensus [1] that age and physical state are the most decisive factors in fall risk assessment.

Applying varying degrees of feature selections on our training data set, we observed that the best results were obtained with the 25 most important features. Even though adding more features was not decreasing the results, it was not benefiting them either and was increasing model's complexity and computing time so a slightly different route was taken from Noh et al.'s work [5]. In our dataset construction, we tried to implement as many information as possible on the older adults as we thought it could lead to better results, however it appeared that there was a balance to have and only a hand few of those information features were actually decisive in the classying process.

This dataset was combined along with algorithmic approaches that were more able to handle high dimensional data, such as Random Forest. Applying data augmentation techniques such as feature interpolation did not improve the results either which is contrary to other similar papers experience with data augmentation such as the work of Yokota et al. [46] who achieved high accuracy detecting falls and augmenting the training samples, and the work of Santos et al. [47] who uses data augmentation to improve its accuracy in the detection of falls. This could be due to the size of our dataset entries, which implied less variations.

In order to optimize the classification learning of the models, we chose to use ensemble methodologies, using Stacking, Bagging and Boosting approaches. Our best results were obtained with Bagging and RF as base model. The cross validated results were quite good compared to the literature with 75.3% AUC (compared to 75.6% on average in similar research) which helps strengthening the approach of using random forest models in the context of medical records based classification problems. Other studies [11], [12] had proven its ability to handle the complexity and high dimensionality of this type of dataset and this study helped confirming this. On the subject of predicting fall time (when the next fall would occur after a fall had been reported for a given older adult), we looked at the individuals that had been classified as future fallers by the models and who actually suffered a fall after the study time.

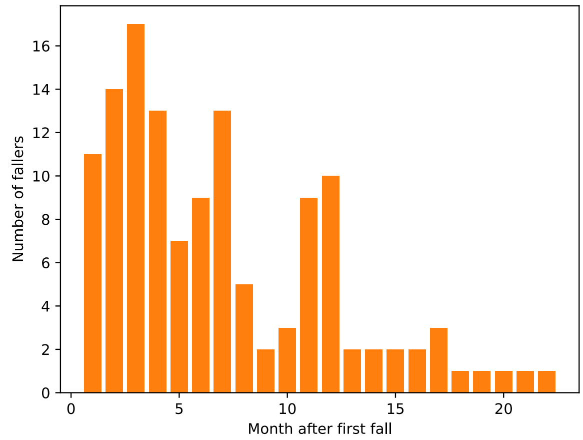


FIGURE 10. **Fallers recorded in the months after first medical visit.**

As can be observed in Figure 10, on average, older adults would be experiencing a fall 7 months after the first clinical visit which gave our model an average prediction time of 7 months. However we still need to take into account that this study is based solely on the data coming from one hospital, the feature extraction applied on the medical records using text mining led to define certain sets of patterns, which led to missing data in certain cases. Data that had to be filled using interpolations and assumptions on the older adults' state. All the models also produced a rather low sensitivity (70% sensitivity, maximum reached for Bagging ensemble with Decision tree base model, which is relatively low compared to an average of 0.847% among similar papers) this is not optimal as it classified a non negligible amount of high risk fallers as non fallers, which in the context of medical prevention, we would rather like to have the contrary (low specificity instead of low sensitivity). Future works should mainly address improvements to make the model dynamic, inputting updated data along the way, in order to better monitor the fall risk index of a given older adult. Having a higher amount of entries in order to analyze their relevance in the context of fall classification in order to pinpoint a coherent list of decisive factors and subsequently for the models to better understand the patterns in high fall risk of older adults. We would like to highlight that collecting data specifically for our experiments could have boosted models' performance since we could have skipped exploiting free comments with potential useful information. However, working with real data supports the validity of our results and their applicability in a real-world scenario. Finally, conducting longitudinal studies to evaluate the long-term effectiveness of the fall risk classification model could provide valuable insights into its clinical utility and impact on patient outcomes.

REFERENCES

[1] *Report on Falls Prevention in Older Age*, World Health Organization, Geneva, Switzerland, 2008.

[2] *Step Safely: Strategies for Preventing and Managing Falls Across the Life-Course*, World Health Organization, Geneva, Switzerland, 2021. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/340962/9789240021914-eng.pdf

[3] P. Rashidi and A. Mihailidis, "A survey on ambient-assisted living tools for older adults," *IEEE J. Biomed. Health Informat.*, vol. 17, no. 3, pp. 579–590, May 2013.

[4] J. Howcroft, J. Kofman, and E. D. Lemaire, "Review of fall risk assessment in geriatric populations using inertial sensors," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, p. 91, 2013, doi: 10.1186/1743-0003-10-91.

[5] H.-M. Noh, H. J. Song, Y. S. Park, J. Han, and Y. K. Roh, "Fall predictors beyond fall risk assessment tool items for acute hospitalized older adults: A matched case–control study," *Sci. Rep.*, vol. 11, no. 1, p. 1503, Jan. 2021, doi: 10.1038/s41598-021-81034-9.

[6] C.-L. Chan, Y.-J. Chen, K.-P. Chen, and S.-J. Chiu, "Elderly inpatient fall risk factors: A study of decision tree and logistic regression," in *Proc. 40th Int. Conf. Comput. Ind. Eng.*, Jul. 2010, pp. 1–6, doi: 10.1109/ICCIE.2010.5668300.

[7] L. Zhang, O. Ma, J. M. Fabre, R. H. Wood, S. U. Garcia, K. M. Ivey, and E. D. McCann, "Classification of older adults with/without a fall history using machine learning methods," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6760–6763, doi: 10.1109/EMBC.2015.7319945.

[8] B. R. Greene, S. J. Redmond, and B. Caulfield, "Fall risk assessment through automatic combination of clinical fall risk factors and body-worn sensor data," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 3, pp. 725–731, May 2017, doi: 10.1109/JBHI.2016.2539098.

[9] C. E. S. Oshiro, T. B. Frankland, A. G. Rosales, N. A. Perrin, C. L. Bell, S. H. Y. Lo, and C. M. Trinacty, "Fall ascertainment and development of a risk prediction model using electronic medical records," *J. Amer. Geriatrics Soc.*, vol. 67, no. 7, pp. 1417–1422, Jul. 2019, doi: 10.1111/jgs.15872.

[10] H. Jung, H.-A. Park, and H. Hwang, "Improving prediction of fall risk using electronic health record data with various types and sources at multiple times," *Comput., Inform., Nursing*, vol. 38, no. 3, pp. 64–157, Mar. 2020, doi: 10.1097/CIN.0000000000000561.

[11] D. S. Lindberg, M. Prosperi, R. I. Bjarnadottir, J. Thomas, M. Crane, Z. Chen, K. Shear, L. M. Solberg, U. A. Snigurska, Y. Wu, Y. Xia, and R. J. Lucero, "Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: A machine-learning approach," *Int. J. Med. Inform.*, vol. 143, Nov. 2020, Art. no. 104272, doi: 10.1016/j.ijmedinf.2020.104272.

[12] C. H. Liu, Y. H. Hu, and Y. H. Lin, "A machine learning-based fall risk assessment model for inpatients," *Comput., Inform., Nursing*, vol. 39, no. 8, pp. 450–459, Aug. 2021, doi: 10.1097/CIN.0000000000000727.

[13] H. G. Hosseini, M. M. Baig, M. J. Connolly, and M. Linden, "A multifactorial falls risk prediction model for hospitalized older adults," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3484–3487, doi: 10.1109/EMBC.2014.6944373.

[14] Y. Lo, S. F. Lynch, R. J. Urbanowicza, R. S. Olson, A. Z. Ritter, C. R. Whitehouse, M. O'Connor, S. K. Keimc, M. McDonald, J. H. Moorea, and K. H. Bowles, "Using machine learning on home health care assessments to predict fall risk," *Stud. Health Technol. Inform.*, vol. 264, pp. 684–688, Aug. 2019, doi: 10.3233/SHTI190310.

[15] A.-H. Tan, "Text Mining: The state of the art and the challenges," in *Proc. Workshop Knowl. Disocovery Adv. Databases (PAKDD)*, vol. 8, 1999, pp. 65–70, doi: 10.1.1.38.7672.

[16] F. N. Patel and N. R. Soni, "Text mining: A brief survey," *Int. J. Adv. Comput. Res.*, vol. 2, no. 6, pp. 243–248, Dec. 2012.

[17] J. Martos. (Sep. 2018). *Make Better Decisions with Text Mining*. [Online]. Available: https://www.techedgegroup.com/es/blog/toma-mejores-decisiones-con-mineria-textos

[18] M. Batta, "Machine learning algorithms—A review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020.

[19] A. L. Hendrich, A. Bufalino, and C. Groves, "Validation of the Hendrich II Fall Risk Model: The imperative to reduce modifiable risk factors," *Appl. Nursing Res.*, vol. 53, Jun. 2020, Art. no. 151243, doi: 10.1016/j.apnr.2020.151243.

[20] I. Bóriková, M. Tomagová, M. Miertová, and K. Žiaková, "Predictive value of the Morse Fall Scale," *Central Eur. J. Nursing Midwifery*, vol. 8, no. 1, pp. 588–595, Mar. 2017, doi: 10.15452/CEJNM.2017.08.0006.

[21] D. Oliver, M. Britton, P. Seed, F. C. Martin, and A. H. Hopper, "Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: Case-control and cohort studies," *Brit. Med. J.*, vol. 315, p. 7115, Oct. 1997, doi: 10.1136/bmj.315.7115.1049.

[22] H. Myers, "Hospital fall risk assessment tools: A critique of the literature," *Inter Nat. J. Nursing Pract.*, vol. 9, no. 4, pp. 223–235, Aug. 2003, doi: 10.1046/j.1440-172x.2003.00430.x.

[23] P. Vega. *Schmid Score Assessment Tool*. SCRIBD. [Online]. Available: https://es.scribd.com/doc/178000152/Schmid-Score-AssessmentTool

[24] JIB Arce. (Jul. 2019). *The Confusion Matrix and Its Metrics*. Health Big Data. [Online]. Available: https://www.juanbarrios.com/la-matrizde-confusion-y-sus-metricas/

[25] L. P. Gómez, S. A. Chicharro, and B. C. Morcuende, "Falls in the institutionalized elderly: Risk assessment, related factors, and description," *Gerokomos*, vol. 28, no. 1, pp. 2–8, Mar. 2017.

[26] H.-M. Hu, "Facing an aging society: Taiwan's universities in crisis," *Gerontol. Geriatrics Educ.*, vol. 41, no. 2, pp. 233–241, Apr./Jun. 2022, doi: 10.1080/02701960.2018.1428576.

[27] G. F. Fuller, "Falls in the elderly," *Amer. Family Physician*, vol. 61, no. 7, pp. 2159–2168, Apr. 2000.

[28] C. H. Torres and T. Hsu, "Comprehensive geriatric assessment in the older adult with cancer: A review," *Eur. Urol. Focus*, vol. 3, nos. 4–5, pp. 330–339, Oct. 2017, doi: 10.1016/j.euf.2017.10.010.

[29] H. Qiu, R. Z. U. Rehman, X. Yu, and S. Xiong, "Application of wearable inertial sensors and a new test battery for distinguishing retrospective fallers from non-fallers among community-dwelling older people," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, Nov. 2018.

[30] N. Zurbuchen, A. Wilde, and P. Bruegger, "A machine learning multi-class approach for fall detection systems based on wearable sensors with a study on sampling rates selection," *Sensors*, vol. 21, no. 3, p. 938, Jan. 2021.

[31] N. A. Schmid, "1989 federal nursing service award winner. Reducing patient falls: A research-based comprehensive fall prevention program," *Mil. Med.*, vol. 155, pp. 202–207, May 1990. [Online]. Available: https://academic.oup.com/milmed/article-abstract/155/5/202/4847473?loginfalse

[32] W.-M. Chu, E. Kristiani, Y.-C. Wang, Y.-R. Lin, S.-Y. Lin, W.-C. Chan, C.-T. Yang, and Y.-T. Tsan, "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," *Frontiers Med.*, vol. 9, Aug. 2022, Art. no. 937216, doi: 10.3389/fmed.2022.937216.

[33] Centers for Disease Control and Prevention. (2022). *Falls and Older Adults*. [Online]. Available: https://www.cdc.gov/homeandrecreationalsafety/falls/adultfalls.html

[34] E. Hossain, R. Rana, N. Higgins, J. Soar, P. D. Barua, A. R. Pisani, and K. Turner, "Natural language processing in electronic health records in relation to healthcare decision-making: A systematic review," *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106649, doi: 10.1016/j.compbiomed.2023.106649.

[35] S. R. Lord, H. B. Menz, and A. Tiedemann, "A physiological profile approach to falls risk assessment and prevention," *J. Amer. Phys. Therapy Assoc.*, vol. 83, no. 3, pp. 237–252, 2003, doi: 10.1093/ptj/83.3.237.

[36] L. Z. Rubenstein, "Falls in older people: Epidemiology, risk factors and strategies for prevention," *Age Ageing*, vol. 35, Suppl. 2, pp. ii37–ii41, 2006, doi: 10.1093/ageing/afl084.

[37] M. Honnibal and I. Montani. (2017). *spaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing*. [Online]. Available: https://spacy.io/usage/spacy-101

[38] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and J. Vanderplas, "*Scikit-learn*: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011. [Online]. Available: https://scikit-learn.org/stable/about.html

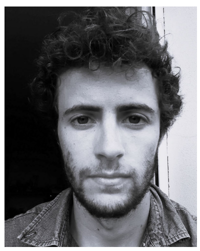
[39] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf

[40] S. Samsir, J. H. P. Sitorus, Zulkifli, Z. Ritonga, F. A. Nasution, and R. Watrianthos, "Comparison of machine learning algorithms for chest X-ray image COVID-19 classification," *J. Phys., Conf. Ser.*, vol. 1933, no. 1, Jun. 2021, Art. no. 012040, doi: 10.1088/1742-6596/1933/1/012040.

[41] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "A review of feature selection methods for machine learning-based disease risk prediction," *Frontiers Bioinf.*, vol. 2, Jun. 2022, Art. no. 927312, doi: 10.3389/fbinf.2022.927312.

[42] D. Yadav and S. Pal, "Prediction of heart disease using feature selection and random forest ensemble method," *Int. J. Pharmaceutical Res.*, vol. 12, no. 4, pp. 56–66, Jun. 2020, doi: 10.31838/ijpr/2020.12.04.013.

[43] A. Acharjee, J. Larkman, Y. Xu, V. R. Cardoso, and G. V. Gkoutos, "A random forest based biomarker discovery and power analysis framework for diagnostics research," *BMC Med. Genomics*, vol. 13, p. 178, Nov. 2020, doi: 10.1186/s12920-020-00826-6.
[44] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4. New York, NY, USA: Springer, 2006.
[45] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc., Ser. B, Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
[46] S. Yokota and K. Ohe, "Construction and evaluation of FiND, a fall risk prediction model of inpatients from nursing data," *Jpn. J. Nursing Sci.*, vol. 13, no. 2, pp. 247–255, Apr. 2016.
[47] G. L. Santos, P. T. Endo, K. H. C. Monteiro, E. D. S. Rocha, I. Silva, and T. Lynn, "Accelerometer-based human fall detection using convolutional neural networks," *Sensors*, vol. 19, no. 7, p. 1644, Apr. 2019, doi: 10.3390/s19071644.
[48] F. M. Martins, V. M. G. Suárez, J. R. V. Flecha, and B. G. López, "Data augmentation effects on highly imbalanced EEG datasets for automatic detection of photoparoxysmal responses," *Sensors*, vol. 23, no. 4, p. 2312, Feb. 2023, doi: 10.3390/s23042312.
[49] J. Fürnkranz, "Decision tree," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA, USA: Springer, 2011, doi: 10.1007/978-0-387-30164-8_204.
[50] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[51] D. W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Hoboken, NJ, USA: Wiley, 2013, doi: 10.1002/9781118548387.
[52] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 3146–3154, doi: 10.5555/3294996.3295070.
[53] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967, doi: 10.1109/TIT.1967.1053964.
[54] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
[55] V. V. Rani, S. Das, and T. K. Kundu, "Risk prediction model for lung cancer disease using machine learning techniques," in *Innovations in Computer Science and Engineering*. Singapore: Springer, Jan. 2022, pp. 417–425, 10.1007/978-981-16-8987-1_44.
[56] K. Makino, S. Lee, S. Bae, I. Chiba, K. Harada, O. Katayama, K. Tomida, M. Morikawa, and H. Shimada, "Simplified decision-tree algorithm to predict falls for community-dwelling older adults," *J. Clin. Med.*, vol. 10, no. 21, p. 5184, Nov. 2021, doi: 10.3390/jcm10215184.
[57] Z. Lyu, L. Wang, X. Gao, and Y. Ma, "The identification of elderly people with high fall risk using machine learning algorithms," *Healthcare*, vol. 11, no. 1, p. 47, Dec. 2022, doi: 10.3390/healthcare11010047.
[58] F. Bu, X.-H. Deng, N.-N. Zhan, H. Cheng, Z.-L. Wang, L. Tang, Y. Zhao, and Q.-Y. Lyu, "Development and validation of a risk prediction model for frailty in patients with diabetes," *BMC Geriatrics*, vol. 23, no. 1, p. 172, Mar. 2023, doi: 10.1186/s12877-023-03823-3.
[59] M. Kim, Y. Lee, J. Lee, J. Lee, S. Lee, and K. Kim, "Comparison of LightGBM and XGBoost algorithms for predicting 30-day hospital readmissions," *IEEE Access*, vol. 8, pp. 103385–103395, 2020, doi: 10.1109/ACCESS.2020.2991385.
[60] J. You, Y.-R. Zhang, H.-F. Wang, M. Yang, J.-F. Feng, J.-T. Yu, and W. Cheng, "Development of a novel dementia risk prediction model in the general population: A large, longitudinal, population-based machine-learning study," *eClinicalMedicine*, vol. 53, Nov. 2022, Art. no. 101665, doi: 10.1016/j.eclinm.2022.101665.
[61] K. G. Gunale and P. Mukherji, "Fall detection using K-Nearest Neighbor classification for patient monitoring," in *Proc. Int. Conf. Inf. Process. (ICIP)*, Pune, India, Dec. 2015, pp. 520–524, doi: 10.1109/INFOP.2015.7489439.
[62] S. Yokota, M. Endo, and K. Ohe, "Establishing a classification system for high fall-risk among inpatients using support vector machines," *Comput., Inform., Nursing*, vol. 35, no. 8, pp. 408–416, Aug. 2017, doi: 10.1097/CIN.0000000000000332.
[63] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements Of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY, USA: Springer, 2009.
[64] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuit Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 3rd Quart., 2006, doi: 10.1109/MCAS.2006.1688199.
[65] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newslett.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.
[66] L. Hwangbo, Y. J. Kang, H. Kwon, J. I. Lee, H.-J. Cho, J.-K. Ko, S. M. Sung, and T. H. Lee, "Stacking ensemble learning model to predict 6-month mortality in ischemic stroke patients," *Sci. Rep.*, vol. 12, no. 1, p. 17389, Oct. 2022, doi: 10.1038/s41598-022-22323-9.
[67] H. V. Nguyen and H. Byeon, "Prediction of Parkinson's disease depression using LIME-based stacking ensemble model," *Mathematics*, vol. 11, no. 3, p. 708, Jan. 2023, doi: 10.3390/math11030708.
[68] A. J. Mahariba, R. AnnieUthra, V. Chandramouleesvar, and M. E. Swetha, "Development of novel Ensembled Boosting Model (EBM) for fall detection," *Adv. Parallel Comput. Algorithms*, vol. 41, pp. 10–18, Nov. 2022, doi: 10.3233/APC220003.
[69] S. H. Lee, J. Lee, and J. Lee, "Fall risk prediction using machine learning in Korean community-dwelling elderly," *Healthcare Inform. Res.*, vol. 25, no. 1, pp. 10–16, 2019, doi: 10.4258/hir.2019.25.1.10.
[70] M. J. Jie et al., "Fall detection system based on inertial sensor and bagging-SVM," *IEEE Access*, vol. 8, pp. 108885–108894, 2020, doi: 10.1109/ACCESS.2020.3008792.
[71] X. Tong, Y. Dong, and M. Xue, "A stacking ensemble learning model for predicting elderly fall risks based on medical records," *IEEE Access*, vol. 9, pp. 139777–139786, 2021, doi: 10.1109/ACCESS.2021.3117760.
[72] A. Madrid and R. Rodriguez, "Application of artificial intelligence techniques for the development of a personalised fall risk index for elderly people," Dept. Comput. Sci., Universidad Carlos III de Madrid, Getafe, Spain, Tech. Rep., 2020.



ANTOINE MILLET received the master's degree (Hons.) in aeronautical engineering from Imperial College London, in 2021. In 2021, he was with the ActiveUP Program, Universidad Politecnica de Madrid, where he is developing a machine learning program to detect and label activities of daily living (ADLs) in older adults living autonomously, to better understand the onsets of frailty. In Fall 2022, he was with the Fundaci para la Investigacin Biomdica HUG, Getafe, Madrid. He started researching fall risk index development for older adults in combination with machine learning methods, leading to the writing of the research paper "Machine learning techniques applied to the development of a fall risk index for older adults," in June 2023.

ALICIA MADRID received the M.Eng. degree in computer science from the University of Carlos III, Madrid, Spain, in 2022. From 2021 to 2022, she was a Research Assistant with the Foundation of Biomedical Investigation, University Hospital of Getafe. She is currently a Software Engineer with the Machine Learning Department, Amazon. Her current research interest includes building predictive models of recurrent falls for older people. The goal of the research was to try and predict if a person was going to have a fall in the next seven months.

JUAN M. ALONSO-WEBER received the Ph.D. degree in computer science. He is currently a Visiting Assistant Professor with the University Carlos III of Madrid, where he specially teaches in the field of compiler designs. He has more than three decades of working experience in the field of neural networks. His current research interests include deep learning applied to image processing, applications of edge AI, and other developments that combine machine learning and electronics.

LEOCADIO RODRÍGUEZ-MAÑAS is currently the Head of the Geriatrics Service of the Getafe University Hospital, Madrid; the Scientific Director of the Spanish Center for Biomedical Research on Frailty and Healthy Aging (CIBERFES), Ministry of Health, and the Ministry of Economy and Competitiveness; and the Co-Director of the Toledo Study on Healthy Aging, carried out in 2,895 elderly people followed over nine years. He has been the Director of the Global Aging Research Network of the International Association of Gerontology and Geriatrics (GARN-IAGG), since July 2017. He has led more than 40 publicly-funded research projects, many of them focused on frailty and functional decline in older people, including the Joint Action on Frailty (ADVANTAGE) funded by DGSANTE-European Union. He has published more than 350 originals in indexed journals peer-reviewed articles and written books and book chapters on various aspects of aging, frailty, and diabetes in older people. He has participated in the preparation of technical reports on these topics for national organizations, such as the Ministry of Health, Scientific Societies, and Research Organizations and Agencies, and for international organizations, such as WHO, PAHO, DG-SANTE, DG-Research, IAGG, Governments of Mexico, and Costa Rica and Chile. He is an International Advisor of WHO and PAHO.

RODRIGO PÉREZ-RODRÁ-GUEZ received the Ph.D. degree in biomedical engineering from Universidad Politcnica de Madrid. Currently, he is a Professor with Universidad Rey Juan Carlos, where he actively collaborates with the Intelligent Robotics Laboratory. He is a senior researcher and a higher education professional. He has also been an eHealth Coordinator with the Biomedical Research Foundation, Getafe University Hospital, and an Adjunct Professor with Universidad Carlos III de Madrid (UC3M). He has participated in many research projects both as a collaborator and the principal investigator. His current research interests include robotics, cognitive models, and monitoring technologies always applied to improve the wellness of the older population.