Learning and Nonlinear Models – Journal of the Brazilian Society on Computational Intelligence (SBIC), Vol. 20, Iss. 2, pp. 47-62, 2022

APPLICATION OF MACHINE LEARNING TOOLS IN THE EVALUATION OF THE RISK OF FALLS IN THE ELDERLY: AN INTEGRATIVE REVIEW

Daniele A. Silva
Federal University of Piauí
Center of Technology
e-mail: d.alves@ufpi.edu.br
Av. Universitária, S/N - Campus Ministro Petrônio Portella
CEP 64049-550, Teresina, PI, Brazil

Nayra F. L. C. Branco
Unified Teaching Center of Piauí
e-mail: nayra.ferreira@ceupi.edu.br
R. Durvalino Couto, 1220
CEP 64049-120, Teresina, PI, Brazil

Hermes Manoel G. C. Branco
Federal University of Piauí
Center of Technology
e-mail: hermescb@ufpi.edu.br
Av. Universitária, S/N - Campus Ministro Petrônio Portella
CEP 64049-550, Teresina, PI, Brazil

Guilherme A. Barreto
Federal University of Ceará
Department Teleinformatics Engineering
e-mail: gbarreto@ufc.br
Av. Mister Hull, S/N - Campus of Pici, Center of Technology
CEP 60455-760, Fortaleza, CE, Brazil

Abstract: This integrative review seeks to present an overview of the application of machine learning (ML) tools in the assessment of the risk of falls in the elderly. We searched the CAPES and IEEE Xplore Periodical databases, articles published in English, Portuguese and Spanish, in the last eleven years. Thirteen articles were selected. Most studies use data from sensors to classify the risk of falling and compare the results obtained with results of clinical tests or history of falls. Some studies carried out the selection of characteristics of the collected signals. Research that compared CI tools and conventional scales pointed to a certain superiority of the former. In the selected articles, Multilayer Perceptron (MLP) neural networks were the most explored. It was possible to observe that the ML tools can be applied in the assessment of the risk of falls in the elderly as a classification resource, showing good results.

Keywords: Artificial intelligence; Elderly Assistance; Falls Accidents; Geriatric Assessment.

1 Introduction

Life expectancy has been restructuring the global demographic profile in recent decades with a progressive increase in the elderly population (Closs and Schwnake, 2012). The gradual fragility of the organism that occurs throughout life favors the development of many clinical conditions in the elderly, however, most deaths from unintentional injury in people over 65 years of age are caused by falls (Kramarow et al, 2015).

According to a 2007 report by the World Health Organization (WHO), approximately 28 to 35% of people over 65 years old suffer falls at least once a year, with an increase of 32 to 42% for people over 70 years old (WHO, 2007). A 2011 cross-sectional study involving participants from 100 municipalities in 23 Brazilian states found a prevalence of falls in the elderly of 27.6% in a one-year period (Siqueira et al, 2011).

Falls negatively affect the quality of life of these people, as in addition to hospitalizations for severe musculoskeletal injuries, they can result in limitations in activities of daily living, loss of autonomy, immobility, loss of confidence when walking, social isolation and depression (Luk et al, 2015; Park, 2017). To implement effective prevention strategies, it is necessary to identify the factors involved and establish the risk of falling again (Luk et al, 2015).

The presence of technology is increasingly evident at different levels of health care. Thus, machine learning (ML) tools have been increasingly explored in studies as a way to automate the classification of elderly people regarding their risk of falling (Lugade et al, 2014; Rivolta et al, 2018). Thus, the objective of this integrative review is to present an overview of the application of ML tools in the assessment of the risk of falls in the elderly.

2 Materials and methods

Among the available research methods, it was decided to use the integrative literature review to work on the topic, so that at the end of the process there is a synthesis of the knowledge already produced in independent studies on the topic (Santos et al, 2012; Souza et al, 2010; Botelho et al, 2011).

To identify suitable studies, a search was carried out in the CAPES and IEEE XPLORE journals databases. The search was carried out from June to October 2021, using combinations of keywords and their synonyms by applying the Boolean operators OR and AND. The keywords used in the search were the following ones: *artificial intelligence, fall risk, elderly, falling, balance, older, sensor and neural network.*

After the initial search, inclusion criteria were applied to refine the search, identifying works published in the last eleven years, in Portuguese, English and Spanish in order to carry out the search encompassing both national and international studies. After filtering, the studies found had their titles and abstracts analyzed and those that dealt with the application of ML tools in the assessment of the risk of falls in the elderly had their full text submitted to methodological quality assessment, according to which they should simultaneously attend to four criteria to be entered in the review. Studies should also report data that assess the performance of the tool, such as sensitivity, specificity and accuracy.

Articles that did not meet the inclusion criteria, in addition to those that were duplicated in the databases, studies unrelated to the theme, editorials and letters to the editor were excluded. Other literature reviews found in the bibliographic research carried out had their reference lists checked in search of more studies that were in accordance with the inclusion and exclusion criteria.

The methodological quality assessment of the articles selected for this stage was performed based on four criteria taken from the guidelines of the Statement on Reporting of Evaluation Studies in Health Informatics (STARE-HI). This tool aims to improve the quality of evaluative studies published in the area of Health Informatics (Talmon et al, 2009). Studies that met the following criteria were included: description of system details, methods for data acquisition, demographic aspects and study results. Such criteria were selected because their presence provides minimal bases for comparing results and reproducing experiments related to the use of ML tools in the context of assessing the risk of falls in the elderly.

3 Results

The search in the databases found 2184 publications. After reading the titles and abstracts, 2158 articles were discarded as they did not fit the proposed theme. The main search left 20 publications suitable for a more complete analysis. Six studies found in the reference lists of four literature reviews were also included, totaling 26 articles for analysis of methodological quality. Thirteen publications were disregarded during this stage, so that thirteen articles were included in this integrative review. The search process followed by this review is represented by the flowchart in Figure 1.

Once identified and evaluated, the articles to be included were listed in chronological order and each one was reviewed in order to extract relevant information. Table 1 contains information from the articles regarding devices or other tools used for data collection, ML tools applied and the size of the sample studied. In Table 2, the methodologies used in each study and the results found by the researchers are detailed. Table 3 contains main ideas from articles.

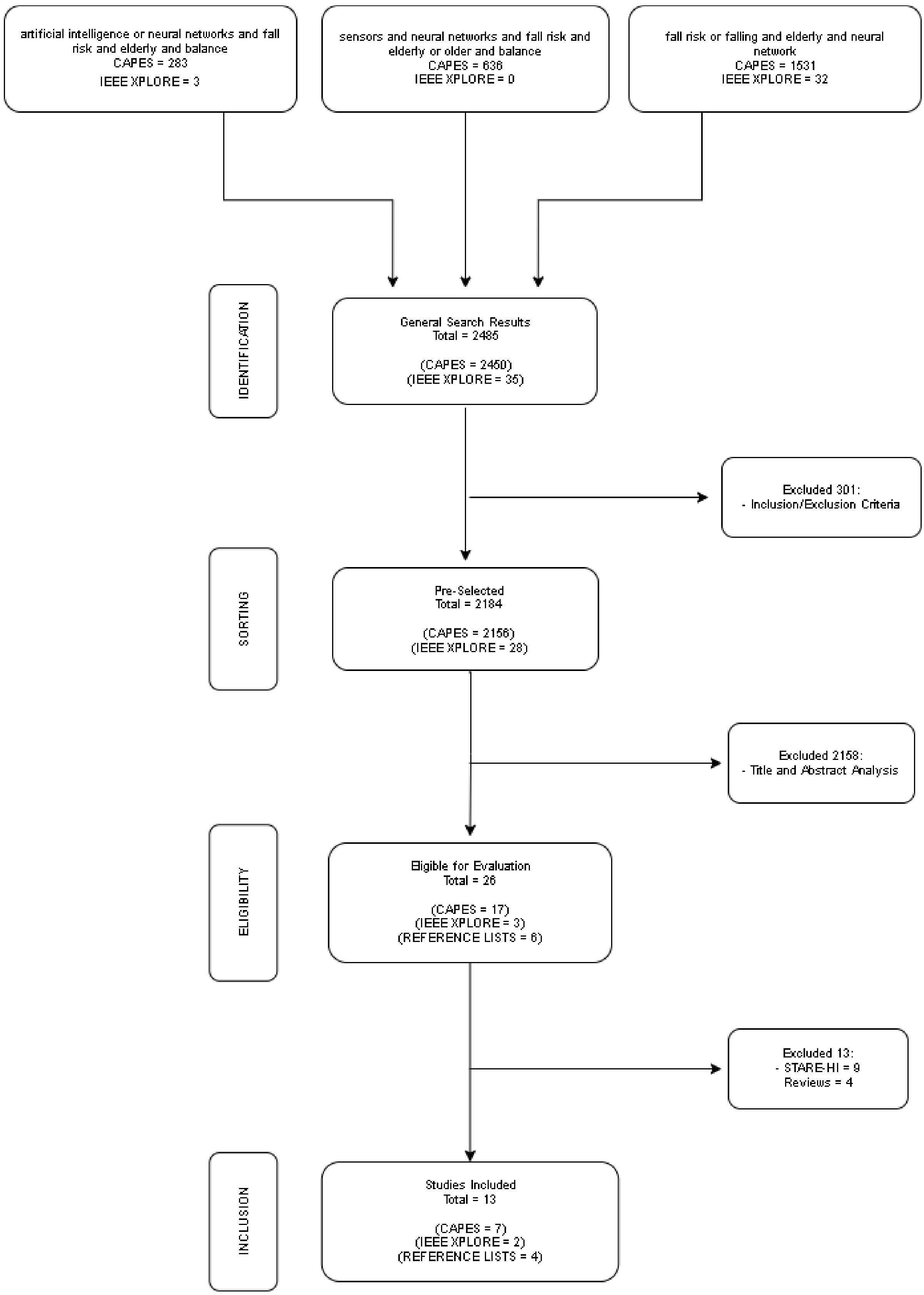

Figure 1: Flowchart of the process of search and selection of studies to be included in this review.
Source: The authors.

Table 1- Characterization of the studies regarding the data acquisition instrument, machine learning tool used and sample size.

Author (year)	Data Acquisition	Machine Learning Tool	Sample size
Buisseret et al (2020)	Triaxial accelerometer, gyroscope and magnetometer fixed to the lumbar region at L4 level.	Convolutional Neural Network (CNN)	73
Caby et al (2011)	Triaxial accelerometers on knees, ankles, elbows, wrists and shoulder blades.	Radial Basis Function Network Classifier (RBNC); Support Vector Classifier (SVC); K-Nearest Neighbour Classifier (KNNC); Naive Bayesian Classifier (NaiveBC)	20
Greene *et al* (2012)	Pressure platform; triaxial accelerometer fixed to the lumbar at the level of the L3 vertebra.	Support Vector Machine (SVM)	120
Howcroft et al. (2016)	Accelerometers attached to the back of the head and pelvis, and side of the leg; plantar pressure sensor.	Multilayer Perceptron (MLP); Naive Bayes (NB); SVM	100
Howcroft et al. (2017a)	Accelerometers attached to the back of the head and pelvis, and side of the leg; plantar pressure sensor.	MLP; NB; SVM	75
Howcroft et al. (2017b)	Accelerometers attached to the back of the head and pelvis, and side of the leg; plantar pressure sensor.	MLP; NB; SVM	100
Lee et al. (2019)	Accelerometer, gyroscope and triaxial magnetometers attached to the head, back, arms, pelvis, thighs, legs and feet.	CNN	50
Marschollek et al (2011a)	Triaxial accelerometer attached to the waist.	Classification and Regression Tree (CART); LR.	46
Marschollek et al (2011b)	Triaxial accelerometer attached to the waist.	LR	46
Razmara, Zaboli and Hassankhani (2016)	Quiz.	MLP	200
Rivolta et al. (2018)	Triaxial accelerometer attached to the chest.	Linear Model (LM); MLP	90
Roshdibenam et al (2021)	Triaxial accelerometer and gyroscope attached to the front of both shoes and the back of the shirt collar.	SVM, CNN	98
Zhang et al. (2015)	3D motion capture system with 35 infrared reflective markers attached to the body; treadmill.	K-Nearest Neighbours (KNN); NB; LR; MLP; SVM	36

Source: The authors.

Table 2 – characterization of studies regarding the methodology applied and results achieved

Author (year)	Methodology	Results
Buisseret et al (2020)	Initially, the participants performed the TUGT (used as the gold standard at this first moment) and soon after, with the sensor fixed to the fourth lumbar vertebra, they performed the 6-minute walk test. Three algorithms were used to classify the elderly. The conventional TUGT was based on time, the TUGT+ used the standard deviation and fractal dimension variability indices and the CNN network with 4 2-dimensional convolutional layers and 3 neural network layers with ReLu function. Six months later, the assessment was redone considering the history of declines in this period as the gold standard.	The TUGT classification obtained an accuracy of 65%, sensitivity of 71% and specificity of 54%. The TUGT+ obtained 73% of accuracy, 85% of sensitivity and 50% of specificity. The CNN network reached 75% in the three metrics. It is noteworthy, however, that the CNN validation dataset had only 16 individuals.
Caby et al (2011)	Participants were evaluated with the Fukuda, Tinetti and Mini Motor tests. With the accelerometers in place, the 25m walk tests, Timed Up and Go (TUGT), physical performance scale and balance on one leg were performed. Sixty-seven features were extracted in the time and frequency domains, to be evaluated using a Wrapper approach of the forward selection type through the RBNC, SVC, KNNC and NaiveBC networks. Some characteristics were modified by subtraction of the mean and analyzed again by the algorithms. The gold standard to be compared with nets was an assessment based on the occurrence of falls and clinical test scores.	With the original features, the NaiveBC obtained the best performances, with 6 of the 7 Baisean algorithms getting 95 and 100% hits, which may suggest overfitting for this technique. The correlation between arm movement was the feature present in all NaiveBC models, as well as features related to frequency (step frequency and peak frequencies) were also present in all other tools, indicating that these features are relevant variables. in the classification of risk of falls. The worst performing tool was the SVC with 75% of correct answers. The modified features did not increase the performance of the tools.
Greene *et al* (2012)	Each participant was evaluated with the BSE and instructed to perform 6 postures, 3 times with eyes open for 40 seconds (OA) and 3 times with eyes closed (OF) for 30 seconds, on a pressure platform with the accelerometer fixed at level of L3. The features derived from the position of the pressure center and the frequency domain extracted were used in 9 SVM models as follows: A) Variables of both sensors based on gender - (M1) men + women, (M2) men and (M3) women; B) Pressure platform variables based on difficulty - (M4) OA + OF, (M6) OA, (M8) OF; C) Accelerometer variables based on difficulty - (M5) OA + OF, (M7) OA and (M9) OF. For the purpose of comparison, BSE data were also used in an SVM model with the direct result of the BSE application, being adopted as the gold standard.	The most effective SVM was model 3 (M3) with an accuracy of 73.33%, using features such as mean distance traveled and mean square root of the Center of Pressure. The model with the lowest accuracy was the SVM 6 (M6) reaching 58.87%, using as features the swing frequency and average swing frequency. The BSE-based model reached 59.42%.
Howcroft et al.(2016)	Plantar sensors were attached to the participants' shoes and accelerometers were attached to the head, waist, and side of the leg. Participants performed a 6m walk under standard conditions and a 7.62m walk with dual task (DT) and single task (ST). Those who reported at least one episode of fall in the six-month period prior to participating in the study were classified as fallers. The plantar sensor allowed the extraction of features derived from the Center of Pressure (COP), such as COP displacement speeds; impulse, derived from the	Under ST the best model was an SVM with accuracy, sensitivity and specificity of 84, 50 and 94.7%, respectively, using features extracted from the plantar and pelvis sensors. The worst model was an MLP based on clinical evaluation with accuracy, sensitivity and specificity of 72, 16.5 and 89.5%, respectively. Under DT the best model was an SVM with accuracy, sensitivity and specificity of 80, 100 and 73.5%, respectively, using

	force-time curve; and time domain, such as cadence and swing time. From the signals obtained with the accelerometer, temporal variables, Fast Fourrier Transform, Ratio of Even to Odd Harmonics and Maximum Lyapunov Exponent were extracted. Combinations of features were then applied to MLP networks, with 5 to 25 hidden neurons, linear and quadratic NB; and SVM. Models based on clinical evaluation were also developed.	features extracted from the plantar and pelvis sensors. The worst model was an NB network with accuracy, sensitivity and specificity of 72, 16.5 and 89.5%, respectively, using only features of the plantar sensor.
Howcroft et al. (2017a)	Plantar sensors were attached to the participants' shoes and accelerometers were attached to the head, waist, and side of the leg. Participants then did a 6m walk under standard conditions and a 7.62m walk with DT and ST. Those who reported at least one episode of fall within six months of participating in the study were classified as fallers. The plantar sensor allowed the extraction of features derived from the Center of Pressure (COP), impulse and time domain. From the signals obtained with the accelerometer, the following were removed: temporal variables, Fast Fourier Transform, Ratio of Even to Odd Harmonics and Maximum Lyapunov Exponent. Combinations of features were then applied to MLP networks, with 5 to 25 hidden neurons, linear and quadratic NB; and SVM. Models based on clinical evaluation were also developed. The twenty best networks were re-evaluated using Repeated Random Sample (RRS) in training.	Under ST, the best model was an SVM with accuracy, sensitivity and specificity of 78.9, 57.1 and 91.7%, respectively, using variables from the sensors of the head and right leg. The worst model obtained accuracy, sensitivity and specificity of 42.9, 72.7 and 50%, respectively, using variables from the sensors of the head, pelvis and both legs. Under DT, the best model was an MLP with accuracy, sensitivity and specificity of 85.7, 72.7 and 66.7%, respectively, using variables from the sensors of the head, pelvis and left leg. The worst model was an NB network with accuracy, sensitivity and specificity of 57.1, 45.5 and 40%, respectively, using variables from the plantar and leg sensors. With the RSS, the best model was an MLP, DT, with an accuracy of 56.5%, using data from the sensors of the head, pelvis and left leg.
Howcroft et al. (2017b)	Plantar sensors were attached to the participants' shoes and accelerometers were attached to the head, waist, and side of the leg. Participants then walked 7.62 m. The plantar sensor allowed the extraction of features derived from the Center of Pressure (COP), impulse and time domain. From the accelerometer signals the following were removed: temporal variables, Fast Fourrier Transform, Ratio of Even to Odd Harmonics and Maximum Lyapunov Exponent. The selection of relevant features among the 146 found was made using 3 methods: correlation-based feature selection (CFS); fast correlation based filter (FCBF) and Relief-F. Combinations of features were then applied to MLP networks, with 5 to 25 neurons in the middle layer; linear and quadratic NB; and SVM. The performance of networks with and without feature selection was then compared. The occurrence of falls in retrospect was adopted as the gold standard. The top ten tools were re-evaluated using Repeated Random Sample (RRS) in training.	The best model was an SVM that used Relief-F for feature selection and achieved an accuracy of 96% with a sensitivity of 100% and specificity of 94.7%. In this model, features of the plantar sensors, such as impulse, and of the head, such as maximums, averages and standard deviations of anterior and posterior accelerations were used. The worst performing model was an SVM that used CFS for feature selection and reached 80% accuracy with 33.3% sensitivity and 94.7% specificity. This model used only the standard deviation of left acceleration provided by the pelvis accelerometer. With the RRS, the best model was an SVM that achieved accuracy, sensitivity and specificity of 77.9, 26.4 and 95.1%, using only the standard deviation of left acceleration provided by the pelvic accelerometer and selected by CFS.
Lee et al. (2019)	The elderly were evaluated with item 14 of the BSE with sensors attached to the body. Six convolutional networks were elaborated: 3 Conv1D and 3 Conv2D. The networks had 3 layers each, and in each layer the size of the convolution kernels was defined as 32, 64	The best network was a Conv2D with 90.68% accuracy using in layers 1, 2 and 3, with convolution kernels of size 32, 32 and 64, respectively. The worst network was Conv1D with an accuracy of 87.76%

	or 128. The Conv1D network extracts the features by traversing the dataset in only one dimension and the Conv2D network traverses the data in two dimensions.	that used in layers 1, 2 and 3, convolution kernels of size 128, 64 and 64, respectively.
Marschollek et al (2011a)	The elderly performed a 20m walk with an accelerometer attached to their waist. A one-year follow-up was carried out in which the elderly had to report episodes of falls and answer the modified Baecke questionnaire, an instrument to assess the level of physical activity. The following variables were extracted: kinetic energy, pelvic balance, mean step duration, step length, number of steps during and standard deviation of gait periodicity, in addition to spectral density distribution parameters such as number of peaks, frequency, width and relative prominence of the first peak and the dominant peak and number of peaks above a relative threshold. Once the characteristics to be used were defined, they were then applied to 4 classifier models: (M1) CRT with sensor data, (M2) LR with sensor data, (M3) CRT with sensor data + Baecke and (M4) LR with sensor data. sensor + Baecke. The occurrence of falls during the follow-up period was adopted as the gold standard for the classification tools.	The CRT model with accelerometer data was 78% accurate with 58% sensitivity and 96% specificity using dominant peak width, number of peaks and dominant peak frequency. Replacing the peak frequency with questionnaire data, the accuracy reached 80% with a sensitivity of 74% and specificity of 82%. The LR models, using the same features as the CRTs, showed lower performance. The tool that used only accelerometer data had an accuracy of 65% with a sensitivity of 42% and specificity of 82%. On the other hand, the LR with data from the questionnaire reached an accuracy of 70% with a sensitivity of 58% and specificity of 78%.
Marschollek et al (2011b)	The elderly were evaluated with TUGT, Barthel index, St. Thomas Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY) and performed a 20m walk with the sensor strapped to their waist. A one-year follow-up was carried out, where the elderly should report episodes of falls, as well as provide information to determine the level of physical activity. The following characteristics were extracted: kinetic energy, pelvic oscillation in the transverse axis, standard deviation of gait periodicity, average step duration, step length, number of steps during TUGT, in addition to spectral density distribution parameters, such as, frequency of the spectral density of the most prominent peak. The characteristics were applied to two LR based models: (M1) clinical data + risk assessment and (M2) accelerometer data + physical activity level. By way of comparison, the efficiency of St. Thomas Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY), TUGT and general cynical assessment, considering the occurrence of falls during the follow-up period as the gold standard.	The STRATIFY scale, TUGT and clinical evaluation showed accuracy of 48, 50 and 55%, respectively. The LR model that used both clinical data and fall risk assessment tools had an accuracy of 72% with a sensitivity of 68% and specificity of 74%. The LR model that used the accelerometer parameters and physical activity level reached an accuracy of 70%, sensitivity of 58% and specificity of 78%.
Razmara, Zaboli and Hassankhani (2016)	Participants were submitted to a questionnaire with items related to activities of daily living (ADLs) and general health conditions. First, the data were applied to two MLP networks: one used data from ADLs, the other used health data. Then the network analysis was redone using Principal Components Analysis (PCA) on the data set to reduce the dimensionality of the features.	Among the networks that did not use the PCA method, the one trained with all the features (ADLs + intrinsic factors) presented the best performance, with an accuracy of 90.2%. The worst performing MLP network without PCA used only the intrinsic factors, reaching an accuracy of 87.5%. The performance of models reduced by PCA was inferior to those that used the feature set in its entirety. The best model reached 89.7% accuracy, considering ADLs and intrinsic factors. The worst performance model

		used only intrinsic factors, obtaining an accuracy of 83.4%.
Rivolta et al. (2018)	Participants were evaluated using the Tinetti scale with an accelerometer attached to the chest. From the evaluation and accelerometer signal, 20 variables related to balance, gait and population characteristics were extracted. After selection with the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm, 9 variables were selected to be applied to the linear model and to the MLP network: duration of the AP triangle, immediate imbalance when standing, ratio between standard deviations of lateral oscillations, sample entropy ratio, step regularity, lateral trunk sway, BMI, gender and age. The MLP network was configured with 5 neurons in the hidden layer and the Tinetti scale was adopted as the gold standard.	Both tools used the 9 selected features. The linear model was 79% accurate with 71% sensitivity and 81% specificity. The MLP network achieved an accuracy of 89% with a sensitivity of 86% and specificity of 90%.
Roshdibenam et al (2021)	The volunteers were evaluated with balance tests such as 30 seconds stand chair and 4 stage balance and TUGT, during which the sensor was active. The collected signal was segmented before being applied to the network, so that the segments of the three signals from each location were used as inputs to the prediction model, a CNN network with 4 layers of 1 dimension, each with Batch normalization and a function of ReLu activation. SVMs were also used for comparison with CNN networks. The dataset was divided so that 80% was destined for the training stage and 20% was directed to the test stage, with the proportion of fallers and non-fallers being maintained equally in both stages. The geriatric assessment, initially used as the gold standard to classify the elderly into fallers and non-fallers, was performed again with the participants after six months. In this last evaluation, the history of falls was used as the gold standard.	In the first evaluation, networks that used data collected from the neck sensor performed better, with the gyroscope data performing slightly better. The CNN network achieved 66% accuracy with 86% sensitivity and 41% specificity. The SVM achieved 67% accuracy, 92% sensitivity and 36% specificity. TUGT obtained 70% accuracy, 56% sensitivity and 88% specificity. At follow-up, SVM achieved 70% accuracy with 2% sensitivity and 96% specificity. The CNN network obtained 60% accuracy, 42% sensitivity and 67% specificity.
Zhang et al. (2015)	The elderly were evaluated using the video system with infrared reflective markers attached to the body. Participants walked on a treadmill at a comfortable walking speed for a maximum of 5 minutes. Spatio-temporal gait variables were extracted from the equipment, such as gait speed, step length and cadence; in addition to kinetic and kinematic characteristics, such as ranges of motion and joint torque. All 32 features were used in the machine learning tools. The following techniques were used: KNN, NB, LR, MLP and SVM. Then, the networks were retrained with PCA-reduced characteristics. Participants who had a history of falls in the last 12 months were considered as fallers.	Initially the best classifier was the MLP network with 77.7% accuracy in the test and the lowest performance tool was the LR with 71.4%. However, it was noticed that most of the tools presented overfitting in the training phase, with the LR reaching 100% accuracy. With the use of PCA, the best tool was the KNN with 85.8% of accuracy in the test and the technique with the lowest performance was the SVM with 73.5%.

Source: The authors.

Table 3 – Main contributions of each article.

Author (year)	**Main Contributions**
Buisseret et al (2020)	The main contribution of this work was to show the good performance of two algorithms for predicting the risk of falls based on sensors, when compared with the performance of the TUG Test, adopting as the gold standard the number of falls recorded 6 months after the tests were carried out. The first algorithm, called TUG+, based on the standard clinical test together with data obtained with sensors during the TUG Test; and the second based only on a convolutional network with sensor data. Additionally, it is highlighted that the signals obtained with sensors were able to point out significant differences between fallers and non-fallers, with the magnitude of anteroposterior acceleration being significantly greater in fallers to reduce the swing phase and shorten unstable periods during gait.
Caby et al (2011)	The authors used an accelerometer network that acquires 3D data from all members and they computed new features in this field of application. An analysis of the various characteristics computed was carried out and it was observed that the correlation between the movement of both arms seems to have a great influence on the imbalance in the elderly, as well as variables related to the step frequency.
Greene *et al* (2012)	The authors use SVM models with data from two types of sensors -a pressure sensitive platform sensor and a body-worn inertial sensor, mounted on the lower back. Features based on RMS values of acceleration and angular velocity were discriminative and therefore included in the models. In addition, the inclusion of data collected with eyes closed and open improved the performance of the models, which suggests that the protocols for assessing the risk of falls in the elderly should include tests in these two conditions. It is noteworthy that it is possible to use both a pressure platform and an inertial sensor to provide data to a SVM model classify the elderly into fallers and non-fallers with similar accuracy.
Howcroft et al. (2016)	This paper presents a comprehensive investigation of fall-risk classification capabilities that included two types of wearable sensors (accelerometers, pressure-sensing insoles), four accelerometer locations (head, pelvis, left and right shank), and three types of models (neural network, support vector machine, naive Bayesian). Furthermore, the effect of cognitive demand on fall risk classification was assessed using single-task (ST) and dual-task (DT) gait. The use of sensors distributed in several segments of the body allowed analyzing which would be more suitable within the suggested methodology. The authors concluded that regardless of the number of sensors used or not together, the algorithms achieved a good classification. However, for a possible quick and low-cost implementation, it is recommended to use only the sensor fixed to the head.
Howcroft et al. (2017a)	This investigation generated elderly fall-risk predictive models based on wearable-sensor-derived gait data and verified that the data collected from the sensor located in the head is useful for processing with ML, as these inputs are related to upper body movement and the management of visual inputs during gait. It was also observed that single-task and dual-task gait assessments provided similar fall-risk model performance.
Howcroft et al. (2017b)	The authors pointed out that fallers have higher posterior acceleration and lower anterior head acceleration when compared to non-fallers, and this feature was important for faller classification. In addition, the authors highlighted the importance of strength and reaction time against the ground during the stance phase of gait for faller classification. Models derived using the reduced feature sets outperformed models derived using the full feature set when classifying fall risk, demonstrating the benefits of feature selection methods when creating faller classification models. Relief-F was able to better handle a dataset with noise and interdependent parameters.

Lee et al. (2019)	The authors concluded that the most frequent body oscillations in the elderly make direct analysis difficult by quantitative values of the sensors, and it is still necessary to evaluate changes in signal behavior over time and the time spent by the elderly during the activity. Two CNN models ware developed - one dimensional and two dimensional - to predict the score of BBS activities using features obtained by the changes of the signals obtained by the sensors. The proposed CNN models performed well and can be used to evaluate fall risk for elderly persons.
Marschollek et al (2011a)	The authors developed a fall risk model based on sensor data that may potentially be measured during typical activities of daily life (a walk, for example), and to evaluate the resulting model with data from a one-year follow-up study. The authors noticed an important characteristic to differentiate fallers from non-fallers: elderly people with a lower risk of falling present greater variability in the time of the stride, in addition to these possibly presenting a higher level of physical activity than the fallers.
Marschollek et al (2011b)	The main contribution of the work was the proposition of two models based on logistic regression, the first using clinical data and traditional tests and the second sensor data obtained from different tests, which obtained similar performance between them and significant improvements in the sensitivity and specificity rates when compared to the results of clinical tests and traditional tools applied in isolation. The authors show that prospective analyzes indicate that the performance of traditional tools (TUG Test, BBE, STRATIFY, Barthel index, among others) have a low success rate. As traditional tools are often subjective, depending on the cutoffs chosen by experts, they may obtain a better hit rate, but even so, they either have low sensitivity or low specificity. The authors also point out that a multidisciplinary analysis by several professionals, and using different traditional tools, provides an improvement in sensitivity and specificity in prospective studies of longer periods, such as 1 year, for example. The proposed models have similar sensitivity and specificity to those obtained by multidisciplinary analyses.
Razmara, Zaboli and Hassankhani (2016)	Psychological factors and the ability to perform simple activities of daily living (ADL) independently, such as: cleaning the house, having meals and going out shopping or visiting a friend, had a great influence on the prediction of the risk of falls in the elderly. The main contribution of this article was to propose an ANN-based approach using levels of ability/independence to perform ADL in association with general health conditions and clinical data (previous illnesses, medication use), collected via a form, as input data.
Rivolta et al. (2018)	As a main contribution, the authors proposed a neural network base model to estimate the Tinetti scale score using balance and gait features extracted from one single triaxial accelerometer. A low entropy detected in the group of fallers suggests an alteration in the neuro-muscular regulation of the standing body balance, regardless of the associated displacement. In addition, attributes extracted during gait were more sensitive in detecting the risk of falling. Another fact to be highlighted is that the non-linear transformation of the sample space resulted in an improvement in the performance of the neural network.
Roshdibenam et al (2021)	This is the first paper to compare a prediction model with a geriatrician's assessment of fall risk, which synthesizes information on fall risk factors (medical health status, gait impairments, and fall history), rather than only relying on the fall incidents or only on the score provided by fall risk assessment instruments, which can increase the error of false negatives. The models that receive as input signals collected by 3 inertial sensors during the TUG test to classify the risk of falls obtained a performance similar to that obtained by a human expert.
Zhang et al. (2015)	Using a 3D motion capture system, significant gait features related to falls risk are extracted. The results show that a dataset with kinetic-kinematic attributes with due pre-processing is useful in classifying the elderly as fallers or non-fallers through machine learning.

Source: The authors.

The Thirteen included studies date from 2011 to 2021, and none of the included articles were published in the years 2013 and 2014. The distribution of publications over the decade is represented by Figure 2. Of the works that met the inclusion criteria, eleven, about 84% of the selected publications were published in journals and the remaining two were presented at conferences.

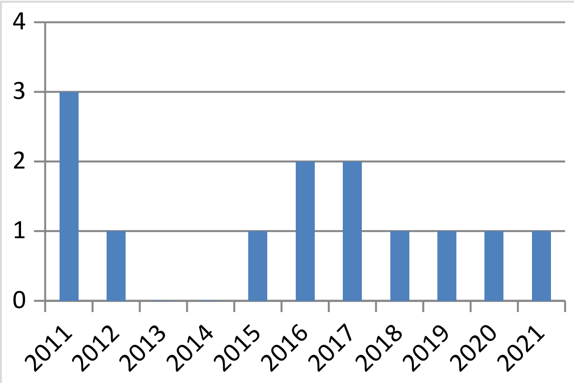

Figure 2 - Distribution of studies in the period between 2011 and 2021.
Source: The authors.

A bibliographic coupling study was performed between the articles included in this review. Bibliographic coupling between two articles occurs when they reference at least one common publication (Grácio, 2016). Therefore, bibliographic coupling measures the degree of connection between two articles by the amount of equal references used in both works. The set of references used by the authors in their articles shows the intellectual environment in which they work and, if two articles present similar bibliographies, there is an implicit relationship between them (Grácio, 2016). The Bibliographic Coupling method is based on the hypothesis that if two articles refer to the same source, they present a similarity, which can be thematic, theoretical, methodological, authors' or other shared particularities. Figure 3 illustrates the bibliographic relationship between studies that analyze the *VosViewer* software. When analyzing this figure, it is observed that the articles can be divided into three clusters, where each one of them retains a certain level of thematic similarity and, possibly, similar lines of research. The first cluster, in green, groups studies that include gait as part of the assessment of the elderly. The second cluster, in red, brings together works that analyzed data from sensors and conventional scales, jointly or not. Finally, the third cluster, in blue, brings together the works of the same group of researchers and naturally have a very similar methodology, using inertial sensors in various parts of the body and plantar sensors.

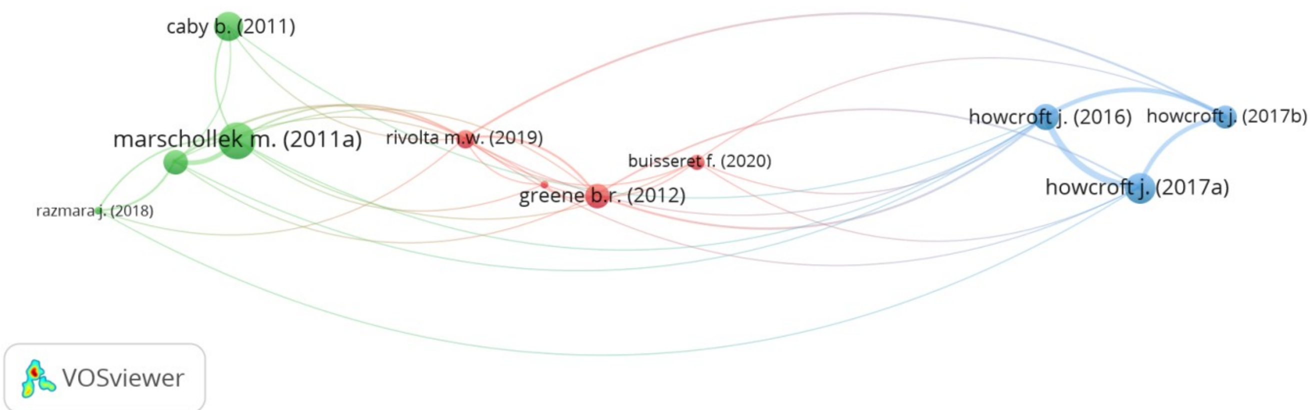

Figure 3 – Bibliographic coupling analysis of the articles founded.

Source: The authors

4 Discussion

With respect to the tools for feature extraction from the data collected, it is noteworthy that, regardless of the sensor used, numerous features can be calculated from the signal provided, both in the time domain and in the frequency domain. In the process of pattern recognition, feature vectors with large dimensions can be harmful to the performance of some architectures of artificial neural networks, since the redundancy of irrelevant parameters increases the computational cost (Lee, 2005), without providing gains related to the percentage of hits. Thus, it is important to extract a large set of features from the signals, however the use of tools for selecting relevant features must be considered and analyzed in the methodologies developed.

The references that specifically paid attention to this point (Caby et al, 2011; Howcroft et al, 2017b), presented two selection approaches that differ from each other by the need or not of a machine learning algorithm in the process. Both were effective, however the wrapper approach (Caby et al, 2011) proved to be slightly inferior to the use of supervised filters (Howcroft et al, 2017b), as overfitting was observed in some networks, a common effect of this method (Loughrey and Cunningham, 2005).

Other researchers have applied feature selection tools in their experiments, despite this not being the main focus of their studies. Rivolta et al. (2018) used the LASSO method to select the features to be used in a linear classifier and an MLP network with 5 neurons in the hidden layer. Zhang et al. (2015) select relevant variables, They used the PCA technique in a study involving several AI tools. In both studies the methods were effective in reducing the parameters with performance improvement in the classifiers. In the study by Razmara, Zaboli and Hassankhani (2016), a neural network of the MLP type was also tested with different configurations of the general dataset and PCA was used to reduce the dimensionality of the set, assembled with data from a questionnaire with information on activities of daily living. and general health conditions. However, unlike Zhang et al. (2015), networks with PCA had a slightly lower performance than those without using the technique, probably due to the nature of the analyzed data and the configuration of the neural network. At this point, it is worth noting that the reduction of dimensionality will not always lead to improvements in classification performance. Often, with the use of a feature vector with lower dimensionality, it is possible to obtain an improvement in computational performance. This aspect should always be carefully taken into account in the development of approaches using computational intelligence methods.

Despite being widely used by health professionals in clinical practice as a prediction tool, the assessment scales are controversial with respect to their real efficiency (Ansai et al, 2014; Oliver et al, 2008). Studies focused on investigating the comparison of computational intelligence against the performance of conventional balance assessment scales achieved satisfactory results using CNN (Buisseret et al, 2020; Roshdibedam et al, 2021), SVM (Greene et al, 2012) and LR (Marschollek et al, 2011a; Marschollek et al, 2011b) models based on data extracted from triaxial sensors attached to the participants' bodies.

It was noted that the performance of the networks was superior to the BSE, TUGT and STRATIFY in classifying the elderly. In fact, in the study by Marschollek et al. (2011b), the performance of the tool, using data from the accelerometer and level of physical activity, was equivalent to the model that used the complete assessment of the multidisciplinary team as input data. This result indicates that the predictive uncertainty of the scales alone is not a parameter to abolish the use of these instruments in the evaluation routines of functional balance, but rather the use of ML tools is to be considered an auxiliary resource in the assessment of the risk of falls in the elderly.

Six articles show comparisons between different ML models in the classification of risk of falls in the elderly. There was a predominance of the use of standard MLP-type neural networks, which in the analyzed studies presented an acceptable performance compared to other types of classification algorithms (Rivolta et al, 2018; Zhang et al, 2015; Howcroft et al, 2017a; Howcroft et al, 2017b). CNN-type deep learning models with different dimensions were also used as an instrument to predict the score of a single BSE item, obtaining similar performances with the different configurations and a slight advantage of the two-dimensional model (Lee et al, 2019).

SVM- and NB-type classifiers were also widely used in studies of comparisons with other networks and with each other, and generally the SVM networks presented a performance close to or superior to the NB (Zhang et al, 2015; Howcroft et al, 2016).

Other types of classifiers were also evaluated. In the study by Zhang et al. (2015), networks with KNN-like classifiers showed a higher accuracy than classifiers, such as SVM, NB and LR. In the study by Marschollek et al. (2011b), logic-based algorithms such as CART were superior to LR models, even with an initial tendency to instability due to a reduced number of participants.

In general, the research analysis also identified the use of specific techniques in order to obtain better accuracy, such as k-fold (Rivolta et al, 2018; Zhang et al, 2015; Greene et al, 2012; Marschollek et al, 2011a; Marschollek et al, 2011b) and leave-one-out (Caby et al, 2011) cross-validation techniques, in addition to the use of specific algorithms for weight initialization, such as the Nguyen-Widrow (Howcroft et al, 2017a). Furthermore, a limitation common to some studies was the use of only the occurrence of falls in a retrospective manner as an algorithm evaluation criterion (Howcroft et al, 2017b; Zhang et al, 2015; Howcroft et al, 2016), since the experience of falling can change the biomechanics of the individual's gait, altering kinetic and kinematics (Kirkwood et al, 2006) later collected with electronic sensors. These changes may even be relevant to favor new episodes of decline. Therefore, an alternative to evaluate the performance of the classifiers is to monitor the volunteers for a period after data collection to verify the occurrence of falls (Caby et al, 2011; Marschollek et al, 2011a; Howcroft et al, 2017a; Marschollek et al, 2011b).

In the selected studies, according to the proposed methodology, there are no studies that use ML tools together with electromyography signals to classify the risk of falls in the elderly. There are a few related works in the literature, but they did not fit the search criteria. It is believed that these signs have great potential to provide relevant information for classifying the risk of falls in the elderly. Electroencephalogram signals may also contain relevant information to classify the risk of falls, however, in the present research, no studies were found that use these types of data as inputs for ML tools.

5 Main challenges for the application of machine learning in the determination of risk of falls in the elderly

Some challenges need to be overcome in the context of using machine learning at risk for the elderly. Below, we highlight the ones that most attract the attention of the authors, based on their experiences and the reports of the articles that composed this systematic review.

Scales and other traditional fall risk analysis tools use the cutoff score to determine the elderly person's degree of risk of falling. However, the cutoff score is not fixed. Studies point to different cutoff scores, regardless of the instrument used, for sedentary and active elderly. Cut-off scores also vary across populations studied, often varying from country to country (Marschollek et al, 2011a). Healthcare professionals can also choose assessments from different cuts, depending on their professional experience. This leads us to believe that trained models to estimate the risk of falls, having as output the classification obtained with the application of a scale, may not respond satisfactorily when applied to a population with different characteristics from the one in the training was performed.

The performances of the fall risk assessment tools, the STRATIFY score, the Timed Up & Go (TUG) test, BBE, Barthel index, among others, are limited. These tools have low sensitivity and specificity when a prospective analysis is performed. The performance may even vary for better or worse, depending on the choice of the cut-off (Thrane et al, 2007; Oliver et al, 2008; Nordin et al, 2008; Kristensen et al 2007; Kim et al 2007).

Still, an elderly person who was indicated by a test with a high risk of falls, but did not suffer a fall, either in a retrospective or prospective analysis, does not necessarily prove that the scale was wrong in pointing out that she had a risk. The fact that patients did not experience a fall does not necessarily indicate that they do not have a gait and balance impairment that might cause a fall. The fall may have been avoided by the care taken. In this sense hand, an elderly person may have suffered a fall episode, even with tests indicating her low risk, and this does not necessarily mean that the method misclassified, as the fall episode may be related to extrinsic factors, for example. (Marschollek et al, 2011b).

This is probably the biggest research challenge on automated ways to determine the risk of falls in the elderly: Defining the best approach or scale, as well as its cut-off scores, to determine the correct exits of what would be a high risk situation or not. In addition, determine a methodology that allows an adequate assessment of whether the performance of the tool was satisfactory or not.

The answer to this challenge may lie in adopting the assessment carried out by a multidisciplinary geriatric team and using different assessment instruments as the desired output. The geriatric care team fall risk score may be perceived as a very subjective measure, yet it represents the professional opinion of several experienced experts that is very likely based on an intuitive understanding of the complex concept 'fall risk' as well as on a multitude of observations of a certain patient.

Other challenges in the area consist of determining which are the best types of sensors to be used, and which tests must be performed to allow acquisition of the signals necessary for a risk assessment. It is worth remembering that the ideal is that the tests are fast, and do not require much effort from the elderly, especially those who are more debilitated. In addition, the chosen sensors, their quantities and locations, should not bring discomfort to the elderly during the tests. It is noteworthy that most works use inertial sensors, but other types of signals, such as dynamometry signals, electromyography and electroencephalogram signals, can provide important information for the assessment of the risk of falls.

Machine learning algorithms that require a large sample size to produce certain results, or those that require multiple attributes, involving multiple types of tests and different sensors, can be a problem, despite providing a reliable answer, especially for clinical evaluations in hospital settings and involving the elderly. These issues are critical because clinicians look for fall-risk detection techniques that are simple and affordable in a clinical setting and can be conducted quickly by a primary doctor. Successful application of feature selection techniques to faller classification would improve the clinical applicability of fall risk prediction models by reducing assessment and analysis complexity. (Roshdibenam et al, 2021)

Finally, it is necessary to improve the engagement of the direct beneficiaries, health professionals, which is still small, probably due to a weak interdisciplinary, from their basic professional training, translated into an innocent ignorance about the possibility of using resources from other areas of knowledge in clinical practice. Linked to this process, there is a need to clarify that the use of technology in the propaedeutic routine will not follow a mechanized route, in order to replace the professional, but will be useful as a complementary tool to the skills of the same.

6 Conclusions

Through the analysis of the articles, it was possible to observe that the machine learning tools are being applied in the assessment of the risk of falls in the elderly as a classification resource. At first, the predominance of comparisons between different models was identified in search of the most efficient technique to classify elderly people with high and low risk.

Secondly, there was a concern to improve the performance of classifiers through data pre-processing techniques. It was observed a wide use of triaxial inertial sensors with fixation in different segments of the body, and the waist region was the preference of most researchers to monitor the variations of displacement of the body's center of mass. Data processing was similar in most of the research, from obtaining characteristics in the time and frequency domain, to filtering the signal to reduce the presence of noise.

Then, comparisons of ML tools with validated conventional scales were identified, and few studies actually carried out a comparative investigation about the effectiveness of these approaches. The main scales used in the studies for this purpose were the Berg Balance Scale and the Timed Up and Go Test. The small amount of research with this purpose exposes the need for new studies of this type, even involving the use of other scales, since it corresponds to an application closer to the reality of care for the elderly.

In general, the ML tools used in the analyzed studies performed well in classifying the elderly as fallers and non-fallers. This promising result will certainly lead to an increase in studies on the classification of risk of fall, since the increase in the number of elderly people tends to increase the demand for specialized care and the use of ML tools can the health care staff to provide adequate assistance to the patients.

Bearing in mind that the assessment of the risk of falling in the elderly must be as stable and efficient as possible, it is necessary to seek to use the minimum number of data recording devices, as well as to investigate whether the use of reduced scales is sufficient. to predict a result that would be obtained with its full version. Furthermore, most of the analyzed studies used kinematic data collected from accelerometers and gyroscopes in inert sensors. Therefore, it is suggested that further research be investigated the feasibility of using electrophysiological signals, such as electrical signals that can be collected by means of electroencephalogram and electromyography.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. H. M. G. C. Branco also thanks CNPq for the grant no. 160132/2018-5. G. A. Barreto also thanks CNPq for the grant no. 309379/2019-9.

7 REFERENCES

Ansai, J. H., Gilsoi, S. F. N., Soares, A. T., Cabral, K. N., Sera, C. T. N. and Paschoal, S. M. P. (2014). Revisão de dois instrumentos clínicos de avaliação para predizer risco de quedas em idosos. Rev Bras Gerontol 17(1):177-189.

Botelho, L. L. R., Cunha, C. C. A. and Macedo, M. (2011) O método da revisão integrativa nos estudos organizacionais. Gestão e Sociedade 11(5): 121-136.

Buisseret, F., Catinus, L., Grenard, R., Jojczyk, L. Fievez, D., Barvaux, V. and Dierick, F. (2020). Timed up and go and six-minute walking tests with wearable inertial sensor: one step further for the prediction of the risk of fall in elderly nursing home people. Sensors 20(11):1-15.

Caby, B., Kieffer, S., Hubert, M. S., Cremer, G. and Macq, B. (2011) Feature extraction and selection for objective gait analysis and fall risk assessment by accelerometry. Biomedical Engineering Online 10(1):1-19.

Closs, E. and Schwnake, C. H. A. (2012). A evolução do índice de envelhecimento no Brasil, nas suas regiões e unidades federativas no período de 1970 a 2010. Rev. bras. geriatr. gerontol 15(3):443-458.

Grácio, M. C. C. (2016). Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. Encontros Bibli: Revista eletrônica De Biblioteconomia E Ciência Da informação, 21(47), 82-99.

Greene, B. R., McGrath, D., Walsh, L., Doheny, E. P., McKeown, D., Garanttini, C., Cunningham, C., Crosby, L., Caulfield, B. and Kenny, R. A. (2012). Quantitative falls risk estimation through multi-sensor assessment of standing balance. Physiol Meas 33:2049–2063.

Howcroft, J., Kofman, J. and Lemaire, E. D. (2017a) Prospective fall-risk prediction models for older adults based on weareable sensors. IEEE Trans Neural Syst Rehabil Eng 25(10): 1812-1820.

Howcroft, J., Kofman, J. and Lemaire, E D. (2017b). Feature selection for elderly faller classification based on wearable sensors. Journal of NeuroEngineering and Rehabilitation 14: 1-11.

Howcroft, J., Lemaire, E. D. and Kofman, J. (2016) Wearable-sensor-based classification models of faller status in older adults. Plos One, pp.1-16.

Kirkwood, R. N., Araújo, P. A. and Dias, C. S. (2006). Biomecânica da marcha em idosos caidores e não caidores: uma revisão da literatura. R Bras Ci e Mov 14(4): 103-110.

Kim EA, Mordiffi SZ, Bee WH, Devi K, Evans D. (2007)Evaluation of three fall-risk assessment tools in an acute care setting. *J Adv Nurs.* **60**:427–435.

Kramarow, E., Chen, L., Heedegaard, H. and Warner, M. (2015). Deaths from unintencional injury among adults aged 65 and over: united states, 2000-2013. NCHS Data Brief 199: 1-7.

Kristensen MT, Foss NB, Kehlet H. (2007) Timed "up & go" test as a predictor of falls within 6 months after hip fracture surgery. *Phys Ther.* **87**:24–30.

Lee, H. D. (2005). Seleção de atributos importantes para a extração de conhecimento de bases de dados. São Carlos: Instituto de Ciências Matemáticas e de Computação da USP.

Lee, Y., Yoon, S. and Kim, W. (2019). A study on CNN-based Berg Balance Scale analysis for elderly persons. Proceedings of 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), pp.1-2.

Loughrey, J. and Cunningham, P. (2005). Overfitting in wrapper-based subset selection: the harder you try the worse it gets.

Trinity College Dublin

Lugade, V., Lin, V., Farley, A. and Chou, L-S. (2014). An Artificial Neural Network Estimation of Gait Balance Control in the Elderly Using Clinical Evaluations. PLoS ONE 9(5):e97595

Luk, J. K. H., Chan, T. Y. and Chan, D. K. Y. (2015). Falls prevention in the elderly: translating evidence into practice. Hong Kong Med J. 21(2): 165-171.

Marschollek, M., Rehwald, A., Wolf, K. H., Gietzelt, M., Nemitz, G., zu Schwabedissen, H. M. and Haux, R. (2011a). Sensor-based fall risk assessment- an expert 'to go'. Methods Inf Med 5:420-426.

Marschollek, M., Rehawald, A., Wolf, K-H., Gietzelt, M., Nemitz, G., Ze Schwabedissen, H. M. and Schulze, M. (2011b). Sensors vs. experts - a performance comparison of sensor-based fall risk assessment vs. conventional assessment in a sample of geriatric patients. BMC Medical Informatics and Decision Making 11:1-7.

Nordin E, Lindelof N, Rosendahl E, Jensen J, Lundin-Olsson L. (2008)Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities. *Age Ageing.* **37**:442–448.

Oliver, D., Papaioannou, A., Giangregorio, L., Thabane, L., Reizgys, K. and Foster, G. (2008). A systematic review and meta-analysis of studies using the STRATIFY tool for prediction of falls in hospital patients: how well does it work. Age Ageing 37(6):621-627.

Park, S. H. (2017). Tools for assessing fall risk in the elderly: a systematic review and metaanalysis. Aging Clinical and Experimental Research 30(1):1-16.

Razmara, J., Zaboli, M. H. and Hassankhani, H. (2016). Elderly fall risk prediction based on a physiological profile approach using artificial neural networks. Health Informatics Journal, pp.1-9.

Rivolta, M. W., Aktaruzzaman, M., Rizzo, G., Lafortuna, C. L., Ferrarin, M., Bovi, G., Bonardi, D. R., Caspani, A. and Sassi, R. (2018). Evaluation of the Tinetti score and fall risk assessment via accelerometry-based movement analysis. Artificial Intelligence in Medicine, pp. 1-10.

Roshdibedam, V., Jogerst, G. J., Butler, N. R. and Baek, S. (2021). Machine learning prediction of fall risk on older adults using timed up and go test kinematics. Sensors 21(10):1-18.

Santos, S. S. C., Silva, M. E., Pinho, L. B., Gautério, D. P., Pelzer, M. T. and Silveira, R.S. (2012). Risco de quedas em idosos: revisão integrativa pelo diagnóstico da North American Nursing Diagnosis Association. Rev Esc Enferm USP 46(5): 1227-1236.

Siqueira, F. V., Facchini, L. A., Silveira, D. S., Piccini, R. X., Tomasi, E., Thumé, E., Silva, S. M. and Dilélio, A. (2011). Prevalence of falls in elderly in Brazil: a countrywide analysis. Cad. Saúde Pública 27(9): 1819-1826.

Souza, M. T., Silva, M. D. and Carvalho, R. (2010) Revisão integrativa: o que é e como fazer. Einstein 8(1): 102-106.

Talmon, J., Ammenwerth, E., Brender, J., Keizer, N., Nykänen, P. and Rigby, M. (2009). STARE-HI - Statement of reporting of evaluation studies in health informatics. International Journal of Medical Informatics 78:1-9.

Thrane G, Joakimsen RM, Thornquist E (2007). The association between timed up and go test and history of falls: the Tromso study. *BMC Geriatr.* **7**:1.

World Health Organization - WHO. (2007). Global report on falls prevention in older age.

Zhang, L., Ma, O., Fabre, J. M., Wood, R. H., Garcia, S. U., Ivey, K. M. and McCann, E. D. (2015). Classification of older adults with/without a fall history using machine learning methods. Proceedings of 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp.6760-6763.