Journal of Population Ageing
https://doi.org/10.1007/s12062-026-09543-x

RESEARCH

The Prediction of Fall Risks in Older Adults: The Analysis of Machine Learning Algorithms and Feature Selections

Etem Curuk[1] · **Merve Pinar**[2] · **Ertugrul Pinar**[3] · **Eyyup Demirpolat**[4]

Received: 22 April 2025 / Accepted: 9 April 2026
© The Author(s) 2026

Abstract

Postural control becomes fragile by aging which increases the risk of falls, leads to fractures, and decreases the quality of life. It is important to investigate the fall risks in people with the older adults. In our study, we analyzed the KINECAL dataset which contains the records of 90 participants aged 18 to 92 years performing clinical movements. We performed six feature selection methods (the Information Gain, Fisher Score (GFS), Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), Chi-square (CS), and Correlation-Based Feature Selection (CBFS)) and four classification algorithms (Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), and Decision Tree (DT)). Our findings showed that XGBoost consistently outperformed the other algorithms, achieving the highest accuracy and F1 scores across most feature selection methods. XGBoost achieved its highest F1 scores with the original dataset (98.28%), CBFS (98.03%), and Information Gain (97.85%). On the other hand, Correlation-Based Feature Selection and Fisher Score generally achieved higher results among the feature selection methods. Conversely, the Support Vector Machine algorithm generally performed poorly. These findings strongly suggest that machine learning methods, particularly ensemble techniques like XGBoost, can be effectively utilized to predict fall risk, and they offer a foundation for developing new approaches in rehabilitation protocols. The study highlights the critical importance of selecting appropriate feature selection methods to improve the performance and interpretability of models used in rehabilitation.

Keywords Postural control · Machine learning · Fall risk detection · Feature selection · Older adults

Extended author information available on the last page of the article

Published online: 06 May 2026

Introduction

Fall is preventable, causes injured-related death and commonly observed in older adults across world. 28%- 32% of the older adults population experience fallin at least one over 65, and the prevelance increases up to 50% over 80 years old population (Organization, 2021). The ability to maintain a balance is crucial to avoid falls during daily activities and deteriorates with aging. The central nervous system (CNS) utilizes sensory information (visual, vestibular, and somatosensory) from the environment to control balance by activating/ inhibiting the muscles. It is necessary that the integration of these sensory inputs is related to body position and sway since sufficient motor commands are required for an upright posture (Peterka, 2002). Older adults have difficulties modulating balance due to reduced sensory inputs and accuracy (Horak et al., 1990) and have at least one fall experience in their lives due to taking some medications, quality of sleep, and the environment (Jahn et al., 2010). In addition, older adults have a higher risk of falls compared to young healthy individuals (Blanchard et al., 2007) and falls are common among the older adults resulting in fractures, disabilities, and increased healthcare expenses (Davis et al., 2010). The ability to maintain balance is well described as associated with the center of mass position (COM) which must be within the base of support (BOS) by activating muscles to maintain equilibrium. The effect of age on postural control results in a larger sway compared to younger adults (Kim et al., 2010). It was suggested that people with older adults demonstrated greater amplitude and velocity of postural sway which resulted the decreased postural stability (Marigold et al., 2004) that may be leading to decreased proprioception, vision, or vestibular sensory inputs (Freeman et al., 2008). It should be noted that the understanding of the multifactorial risk factors that contribute to fall risks in older adults could help evidence-based prevention protocols. The fall prediction models based on medical and monitoring records, emotional and physical assessments have been used in order to classify of older adults as faller and non- fallers. Therefore, falls often involve with internal and external factors (Millet et al., 2023). For example, some studies identified many risk factors in the older adults such as depression dizziness, cognitive impairments, metabolic diseases, sleep problems, hearing and vision impairments. However, the most common reason for the increased risk fall was impaired balance and gait (Fuller, 2000). Although many physical or questionnare- based measurements have been improved to predict or assess fall risks in older adults, there are six tests (Timed Up and Go, Berg Balance Test, Functional Reach Test, Gait Speed Test etc.) commoly used balance test was reported (Myers, 2003). However, this study found that Area under the curve (AUC) for mobility assessments and fall history were between 0.5–0.7 with specificity and sensitivity (ie: TUG specificty: 28.4%- 96.6%, sensitivity: 10%- 83.3%) which provides insufficient consistency (Myers, 2003).

In the literature, different methods has been used for fall prediction. For example, a model was proposed based on statistical methods using data from 204 older adults with similar ages divided into two groups as suffered falls and non-suffered falls (Schmid, 1989). Another relevant approcah was conducted using variable analysis (based on statistical methodology) seperated to experienced fall and non-experienced fall (Bóriková et al., 2017). Oliver et al. (1997) improved the model known

as St. Thomas risk assessment and become commonly in clinical settings. All taken together, different approaches were applied regarding the selection of variables but majority of studies suggested that the balance, gait and mental state are common risk factors (Hendrich et al., 2020). Since variety of variables are available in the clinical research for traditional data analysis, the traditional data analysis for prediction of fall risks in older adults has been challenging. Therefore, traditional way of data analysis is very sensitive with multidimetional datasets, causing the increased risk of false conclusions (Torres & Hsu, 2017). Recently, artificial intelligence (AI) has rapidly emerged as a transformative force in clinical settings by overcoming those challenges using cross-validations to evaluate the results. Many studies used the wearable sensors for data processing for classifying the risk of falls in older adults (Qiu et al., 2018; Zurbuchen et al., 2021). For example, inertial measurement units (IMUs) which can be placed various location of body parts are used for acquisition and analysis (Ferreira et al., 2022). However, its application to postural control and fall risk stratification remains underexplored. Although prior studies have examined the association between biomechanical sway parameters and clinical assessments, and some have applied statistical models or machine learning techniques to fall prediction, these studies balance parameters and do not provide performance of models or detailed information (Arnold et al., 2005; Howcroft et al., 2013). For example, Zhang et al. (2015) used an algorithms to investigate the risk falls with relatively small sample ($n=35$) over 60 years old by applying K- Neural Network, Naives Bayes, Logistic Regression and Supprot Vector Machine including tenfold- cross validation for predicting a first fall within 12 months. Another study investigated the fall risk using a systematic comparison multifactorial assessments with spatiotemporal parameters based on machine learning methods in people with community-dwelling older adults (Silva et al., 2017). In general, previous studies indicate that impairments of gait and balance are associated with an increased risk of falls. In our study, we aimed to address a common limitation in studies on fall risk prediction. Rather than focusing on a single machine learning model, we designed a comprehensive experiment to explore the interplay between different algorithms and feature selection techniques with four algorithms and six different feature selection methods. Therefore, we gained a deeper understanding of not just performance, but also the stability and real-world generalizability of the models. To ensure the trustworthiness and clinical applicability of our results, we took two critical steps: analyzing the data for any age-related bias and using a careful validation strategy to avoid the common pitfall of data leakage. To achieve that, we have an access to KINECAL dataset (https://physionet.org/content/kinecal/1.0.3/) to examine the predictive capacity of machine learning models under different feature selection paradigms. We hypothesize that a systematic machine learning approach can significantly advance the prediction of fall risks. Specifically, we propose that integrating diverse postural sway characteristics, obtained under various body configurations, with relevant clinical assessment variables from dataset. Furthermore, through a comprehensive and transparent evaluation of the interactions between multiple feature selection strategies and diverse machine learning classifiers, this study aims to reveal stable, robust, and clinically meaningful indicators of fall risk. Eventually, the main contributions of this study are as follows:

- This study provides a systematic evaluation of the interaction between multiple feature selection strategies and diverse machine learning classifiers for multi-class fall risk prediction using the KINECAL dataset (Fig. 1).
- It investigates how different dimensionality reduction approaches influence classifier performance, predictive stability, and generalization behavior rather than relying on a single fixed modeling pipeline.
- It examines the consistency of selected predictors across feature selection paradigms in order to identify clinically meaningful and robust indicators associated with fall risk in older adults.
- It integrates class imbalance handling within the cross-validation framework to ensure unbiased performance estimation and to prevent data leakage during model evaluation.
- It advances methodological transparency in AI-driven fall risk modeling by emphasizing both predictive performance and interpretability within a structured comparative framework (Fig. 2).

Methods

Dataset

The dataset was created by the KINECAL (Maudsley-Barton & Yap, 2023) and contains records of 90 participants performing 11 clinical movements commonly used to assess balance impairment and fall risk. The age range of participants was from 18 to 92 years and they were grouped in terms of age and fall risk. The KINECAL

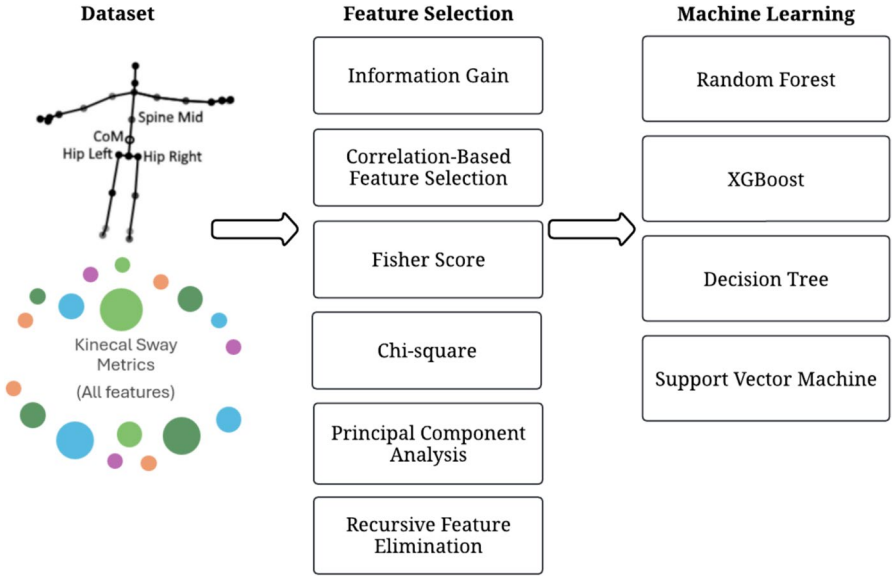


Fig. 1 Flow diagram of fall prediction with KINECAL sway metrics

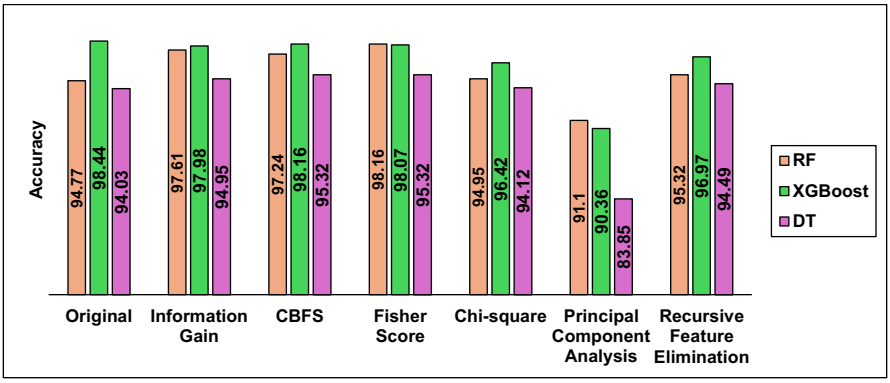


Fig. 2 Figure shows the comparison of accuracies of feature selections and the algorithms of classification. X axis is ordered as original, information gain, CBFS, Fisher Score, Chi- Square, Principal Component Analysis, Recursive Feature Elimination. Y axis is Accuracy in percentile. RF, XGBoost, and DT represent Random Forest, eXtreme Gradient Boosting, Decision Tree, respectively

dataset includes clinical measurements such as Sit Up × 5 (STS-5), 3 m walk, and various postural parameters. The demographics of participants were presented as age, height, weight, and clinical risk labeling. The records were reviewed for quality and problematic data were removed if necessary. The dataset includes sway metrics obtained through the records and these metrics can be used to assess postural control. The KINECAL dataset includes clinically relevant assessments and conditions performed by the participants, such as standing with eyes open and closed, semi-tandem stance, tandem stance, 3 m walk, and sitting and standing 5 times. In the dataset, all participants were asked whether they had experienced a fall in the last 12 months. The response options were labeled "None," "One," "Two," "Three," and "Four or more". Participants were divided into three groups regarding their answers and age: (1) The "Healthy Adult" group consists of those under 65 years of age who have never experienced a fall. (2) The "Non-faller" group consists of participants over the age of 65 who have not experienced a fall. (3) The "self-reported fallers" group consists of participants over the age of 65 who have experienced one or more falls. Participants who reported themselves as self-reported fallers can be divided into two groups "single fallers" and "multiple fallers". "Single fallers" are those who answered "One" to the fall history question and may have a low risk of future falls, while "multiple fallers" are those who have experienced more than one fall in the past and are more likely to fall in the future. This categorization could provide the flexibility to better understanding of assessment of fall risk. The groups in the dataset are shown in Table 1. 521 data samples were acquired. In this dataset, there are 218 records for the HA label, 187 records for the NF label, 80 records for the FHs label, and 36 records for the FHm label. The features of the dataset are shown in Table 2.

The number of labels in the dataset appears to be unbalanced which could adversely affect the performance of the classification model and cause difficulties in learning minority class. The SMOTE (Synthetic Minority Over-sampling Technique) method is utilized to overcome this issue.

Table 1 Distribution of participant groups by age and fall history

Group	Description	Mean age	Participants (*n*)	Gender (f/m)
<65 years, Healthy-Adults	no history of falls in the last 12 months	46.2	33	22/11
Non-Fallerss no history of falls in the last 12 months	73.3	≥65 years, 33	16/17	
Self-reported-single_Faller in the last 12 months	≥65 years, reported 1 fall,	72.3	15	7/8
Self-reported-multiple_Faller	≥65 years, reported>1 fall, in last 12 months	73.1	9	2/7

SMOTE

The advanteages of artificial intelligence were introduced for precision and accuracy of systems for medical field to deal with sensitivity of dataset (Ahmed et al., 2020). In clinical research, SMOTE was used for different purposes. SMOTE is a method used in machine learning to overcome the problem of imbalanced datasets. Imbalanced datasets refer to datasets where one type of data significantly oversamples the other class which creates biases for model performance SMOTE provides a function by generating synthetic examples for the minority class to balance the dataset (). As a data preprocessing method for classification algorithms, SMOTE has been shown to improve the performance of numerous classifiers. It has been shown that rebalancing the class distribution in the training data enables classifiers to effectively learn the data leading to a better prediction performance. For example, imbalanced data was preprocessed using SMOTE and then was fed with five classifiers (Bagging, Support Vector Machine, Multi-Layer Perceptron, Simple Logistic, and Decision Tree) in order to select the best appropriate classifier for a balanced dataset for predicition of diabetes (Shuja et al., 2019). The principles of SMOTE involve identifying the k-nearest neighbors of an instance from the minority class in the feature space. By creating synthetic data points along the lines connecting these neighbors, it generates new synthetic instances for the class with a small number of instances (Fernández et al., 2018). To address class imbalance, the standard Synthetic Minority Oversampling Technique (SMOTE) was employed using a k-nearest neighbor interpolation strategy ($k=5$). The dataset was first divided into training and test subsets, and SMOTE was applied exclusively to the training data to prevent information leakage. The test set remained completely untouched during synthetic sample generation to ensure unbiased performance evaluation.

After oversampling, the augmented training set consisted of 1,090 samples, including 218 FHm, 218 FHs, 218 HA, and 436 NF instances. By increasing the representation of minority classes, SMOTE facilitated more stable model learning and reduced the adverse impact of class imbalance without altering the independent evaluation

Table 2 Original features of the Kinecal dataset and features selected based on feature selection methods

Original	Information gain	Correlation based	Fisher score	Chi-square	Recursive feature elimination	
BMI	MDIST_AP	BMI	Age	Age	Records	Age
Weight	MVELO	Weight	Records	Records	Age	Height
Age	TOTEX_ML	Age	Sex	Sex	Risk	RDIST_ML
Height	MFREQ	Height	Risk	BMI	Sex	RDIST_AP
MFREQ_ML	TOTEX	MFREQ_ML	Height	Height	Movement	RDIST
MVELO_AP	RDIST	MVELO_AP	BMI	Risk	RDIST_AP	MDIST_AP
RDIST_AP	MDIST_ML	RDIST_AP	RDIST_AP	Movement	MDIST_AP	MDIST
MFREQ_AP	AREA_CE	MFREQ_AP	Movement	Weight	BMI	TOTEX_AP
Records	MDIST	Records	MDIST_AP	RDIST_AP	MVELO_AP	MFREQ_AP
Sex	Movement	Sex	MFREQ_AP	MDIST_AP	MFREQ_ML	AREA_CE
MVELO_ML			RDIST_ML			
TOTEX_AP			Risk			

Descriptions of Table 2

Table 2 summarizes the sway metrics obtained from the KINECAL dataset. RD denotes Resultant Distance, AP denotes the anterior–posterior direction, and ML denotes the medio–lateral direction

MDIST Mean distance of the RD time series

RDIST RMS distance of the RD time series

MVELO Mean velocity of the RD time series

TOTEX Total excursions of the RD time series

MFREQ Mean frequency of the RD time series

MDIST_AP Mean distance of the AP time series

RDIST_AP RMS distance of the AP time series

MVELO_AP Mean velocity of the AP time series

TOTEX_AP Total excursions of the AP time series

MFREQ_AP Mean frequency of the AP time series

MDIST_ML Mean distance of the ML time series

RDIST_ML RMS distance of the ML time series

MVELO_ML Mean velocity of the ML time series

TOTEX_ML Total excursions of the ML time series

MFREQ_ML Mean frequency of the ML time series

AREA_CE 95% Confidence Area

protocol. As SMOTE generates new samples through interpolation between existing minority class instances, the synthetic data remain confined within the observed feature space. Since the variables in the dataset correspond to physiologically bounded clinical and postural measurements, the generated samples are expected to remain within realistic and plausible ranges.

Feature Selection

Feature selection is a machine learning process that selects the most relevant features from a dataset to enhance model performance, reduce overfitting, and improve interpretability. This method supports the elimination of irrelevant features, improves prediction accuracy, increases model training and prediction speed, and reduces computational complexity (Saeys et al., 2007). Table 2 presents the original features of the KINECAL dataset and the features obtained using six different feature selection methods. It was ordered by the highest gaining features.

Beyond predictive performance, Table 2 reveals a clear convergence pattern across feature selection methods. Age was consistently selected by nearly all filtering and wrapper-based approaches (CBFS, Fisher Score, Chi-square, and RFE), highlighting its structural contribution to group separation. However, age was not selected in isolation. Biomechanical sway-related metrics such as RDIST_AP, MDIST_AP, MFREQ_AP, and MVELO_AP were repeatedly retained across multiple methods, indicating that postural variability parameters play a central role in discriminating fall-risk categories.

Additionally, clinically contextual variables including Records and Risk were selected by several independent methods, suggesting that fall history and clinical risk labeling provide complementary discriminative information. The simultaneous recurrence of demographic (Age, BMI, Height), clinical (Risk, Records), and dynamic sway-based parameters demonstrates that the model behavior is not driven by a single dominant predictor but rather by a multi-dimensional interaction structure.

Information Gain

Information gain (IG) in feature selection evaluates the information content of features to determine their importance in classification (Jadhav et al., 2018). This method measures the contribution of each feature to classification accuracy, helping to identify the most informative features for effective classification. The IG formula is based on entropy and information theory.

$$\mathrm{IG}\,(\mathrm{D}, \mathrm{A}) = \mathrm{H}\,(\mathrm{D}) = \mathrm{H}\,(\mathrm{D}|\mathrm{A}) \tag{1}$$

IG(D, A) represents the Information Gain of feature A in dataset D (1). H(D) is the entropy of the dataset D before any feature selection is made. H(D|A) is the conditional entropy of dataset D after partitioning the dataset using feature A.

Fisher Score

The Fisher Score (FS) is a supervised feature selection method that evaluates how well features distinguish between classes by considering within-class and between-class variance (Gu et al., 1202). A high Fisher Score means a feature effectively discriminates between classes, making such features preferable for improving model performance.

$$\text{Fisher Score} = \frac{(\mu 1 - \mu 2)^2}{\sigma 1^2 + \sigma 2^2} \tag{2}$$

In the formula (2) μ1 is the mean value of the first class, μ2 is the mean value of the second class, σ1 is the variance of the first class and σ2 is the variance of the second class. Ratios are the difference between the means of two classes to the sum of the within-class variances. In this way, the features with high discrimination power between classes are identified. Features with a high Fisher Score have the potential to discriminate well between classes.

Chi-square

The chi-square (CS) test is a hypothesis testing method used on discrete data to determine whether two variables are dependent or independent (Zhai et al., 2018). This method has two steps: (1) the chi-square statistics of the features by class are calculated. (2) inconsistent features in the data set are separated by applying the chi-merged principle according to the degrees of freedom and the determined significance level. In this process, a higher chi-square value indicates that the relevant factor is more descriptive for the dataset (Gárate-Escamila et al., 2020).

Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique commonly used for dimension reduction and data analysis that transforms original variables into new uncorrelated variables defined as principal components through linear combinations. This technique works by identifying feature sets to simplify complex data sets. PCA aims to capture the maximum variance and minimize the loss of information by searching for linear combinations of the original variables. PCA is another effective method for feature selection in multidimensional data sets (Song et al., 2010).

Recursive Feature Elimination

Recursive Feature Elimination (RFE) is a method that iteratively selects the most important features by training a model on the full feature set and systematically removing the least important ones. This process ranks features based on their predictive importance, reduces overfitting, and enhances model efficiency, helping to identify the most informative features for accurate predictions (Guyon et al., 2002).

Correlation-based Feature Selection (CBFS)

CBFS is a method that uses correlations between features and target classes to identify the most important features of a model (Hall, 1999). The method focuses on selecting features that show strong relationships with the target class while minimizing the correlations between features. By evaluating the correlations between features and the target class, CBFS helps to identify the most meaningful feature subsets

for prediction. This ensures that the most valuable data is selected to improve the performance of the model.

Classification Algorithms

Random Forest (RF)

RF is a powerful statistical classifier that is remarkably fit for its flexibility and high accuracy in forecasting issues (Breiman, 2001). This algorithm is an ensemble method that aims to predict by combining the results of multiple independent decision trees. Random Forests is based on decision trees and ensemble methods (Rigatti, 2017). The algorithm does not rely on assumptions of normality or independence of data, therefore it could work on both continuous and categorical variables and is also robust to outliers, noise, and autocorrelation (Biau & Scornet, 2016).

Support Vector Machine (SVM)

SVM is a supervised learning algorithm used for classification tasks. It identifies the optimal hyperplane that separates different classes by maximizing the margin between them. Originally developed for two-class linear data, SVM has been extended to handle multi-class and non-linear datasets. The points that define the margin are called support vectors and are crucial in determining the hyperplane's position (Wang & Chen, 2020).

eXtreme Gradient Boosting (XGBoost)

XGBoost, or eXtreme Gradient Boosting, is an ensemble learning method that enhances model performance by combining multiple weak learners. It iteratively fits new models to the residuals of previous models, reducing errors and optimizing predictions. By using gradient boosting, XGBoost creates a strong, accurate prediction model from several low-accuracy trees, making it highly effective for various datasets in modern machine learning (Friedman et al., 2000).

Decision Tree (DT)

DT is machine learning and decision analysis tools that represent decision-making processes with multiple stages. They use nodes for decision points, branches for possible outcomes, and leaves for final decisions. Decision trees simplify complex decision-making by breaking it down into sequential steps and ensuring that decisions align with the overall goal, helping decision-makers maintain clarity and consistency (Pinar et al., 2025). The baseline hyperparameter configurations used in this study are presented in Table 3.

All experiments were conducted using Python (version 3.12). Machine learning models were implemented using Scikit-learn (v1.6.1) and XGBoost (v3.2.0). Data processing and analysis were performed using NumPy (v2.1.3), Pandas (v2.2.3), and SciPy (v1.15.2). Visualizations were generated using Matplotlib (v3.10.1) and

Table 3 Baseline hyperparameter configuration

Model	Hyperparameters
Decision Tree	criterion = entropy
Random Forest	n_estimators = 100
Logistic Regression	penalty = L2; C = 1.0; max_iter = 1000
SVM	kernel = linear; C = 1.0

s

Seaborn (v0.13.2). The experimental pipeline was executed in a Windows-based environment.

Evaluation Metrics

Evaluation metrics in machine learning are various metrics used to measure the performance of classification models. These metrics include concepts such as True Positive (TP), True Negative (TN), False Positive(FP), False Negative(FN), classification accuracy, precision, recall, and F1-Score. They show how well the model generalizes on training data and how successful it can be with new data. They also help to determine which algorithm is more effective by evaluating the algorithms' ability to predict availability problems. As a result, these metrics objectively evaluate the performance of the model and draw a roadmap for improvement (Zhou et al., 2021).

The accuracy metric (3) is a widely used measure in machine learning that calculates the ratio of correctly classified instances to total instances. However, the accuracy metric may not provide a complete assessment of the performance of the model, especially in the case of imbalanced data sets. In addition, there are other important metrics such as precision, recall, and F1 score in evaluating classification models. These metrics help to understand the model performance in more detail.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

Precision (4), recall (5), and F1 score (6) are the key metrics for evaluating the performance of machine learning models, especially in classification tasks. Precision is the ratio of correctly predicted positive examples to all positively predicted examples. Recall measures the ratio of correctly predicted positive samples to true positive samples. The F1 score is the harmonic mean of precision and recall. Thus, it provides a balanced assessment of model performance, taking into account false positives and false negatives.

Results

In our study, six different feature selection methods were applied and four different classification results were compared using the original dataset and the features obtained as a result of feature selection. The dataset used in the study is the Kinecal sway metrics dataset and there are 26 features in this dataset. Among these features, "groups" was determined as the target variable, and the variable "part_id" was removed from the dataset since it did not contain any information. Therefore, the experiments were carried out with 24 features in the original dataset.

Using feature selection algorithms, the number of features, which was initially 24, was reduced to 10 and the results obtained were analyzed in detail. Among the feature selection methods, IG, CBFS, FS, CS, PCA, and RFE methods were used. Table 2 shows the features selected using the feature selection methods specified with the original dataset. When the selected features are analyzed, it is observed that "age", "BMI", "sex", "movement", "records", "height", and "risk" have a high effect in common. The k-fold cross-validation method was used for training and testing the model. In this method, the data set is divided into five equal parts (folds) and each is used as a test set in turn, while the remaining four parts are used for training. This process was repeated five times to evaluate the performance of the model more reliably.

Experiments were implemented on four different machine learning algorithms using the features obtained by feature selection methods: RF, SVM, XGBoost, and DT. Each algorithm was trained and tested using feature sets reduced with the specified feature selection methods. The model performances were analyzed using commonly used evaluation metrics such as accuracy, precision, recall, and F1 score.

Within the scope of cross-validation, performance metrics were calculated separately for each fold, and the results were reported as the average metric values across all folds. For example, for the XGBoost model, accuracy was found to be 98.44%±1.31, and the F1-macro score was 98.28%±1.56. Low standard deviation values indicate that the model performance is consistent across different data splits.

Tables 4 and 5 present the comparative performance of classification algorithms across different feature selection strategies. XGBoost consistently achieved the highest scores across accuracy, precision, recall, and F1 metrics, suggesting that ensemble-based boosting methods are particularly effective in capturing the complex structure of postural sway and clinical features. Random Forest also demonstrated

Table 4 Accuracy and precision results of feature selection methods according to classification algorithms

Accuracy				Precision				
RF	SVM	XGB	DT	RF	SVM	XGB	DT	
Original	94.77	70.00	**98.44**	94.03	95.37	68.86	**98.23**	93.71
Information Gain	97.61	63.48	**97.98**	94.95	97.85	75.65	**97.88**	94.75
CBFS	97.24	62.56	**98.16**	95.32	97.25	49.14	**97.96**	94.96
Fisher Score	**98.16**	62.75	98.07	95.32	**98.06**	49.77	97.85	94.96
Chi-square	94.95	62.29	**96.42**	94.12	95.07	48.81	**96.18**	93.85
Principal Component Analysis	**91.10**	67.06	90.36	83.85	**91.95**	80.03	90.54	83.95
Recursive Feature Elimination	95.32	67.06	**96.97**	94.49	96.02	72.11	**96.90**	94.46

Table 5 Recall and F1 score results of feature selection methods according to classification algorithms

Recall				F1 Score				
RF	SVM	XGB	DT	RF	SVM	XGB	DT	
Original	94.26	65.53	**98.33**	93.92	94.73	62.44	**98.28**	93.77
Information Gain	97.30	56.02	**97.82**	94.95	97.55	53.21	**97.85**	94.82
CBFS	97.01	55.10	**98.10**	95.52	97.12	50.68	**98.03**	95.20
Fisher Score	**98.10**	55.10	98.05	95.12	**98.08**	50.72	97.94	94.94
Chi-square	94.83	54.93	**96.38**	93.97	94.92	50.51	**96.28**	93.88
Principal Component Analysis	**90.08**	59.28	89.56	83.31	**90.88**	60.86	90.00	83.53
Recursive Feature Elimination	94.66	61.52	**96.73**	94.38	95.28	59.85	**96.81**	94.41

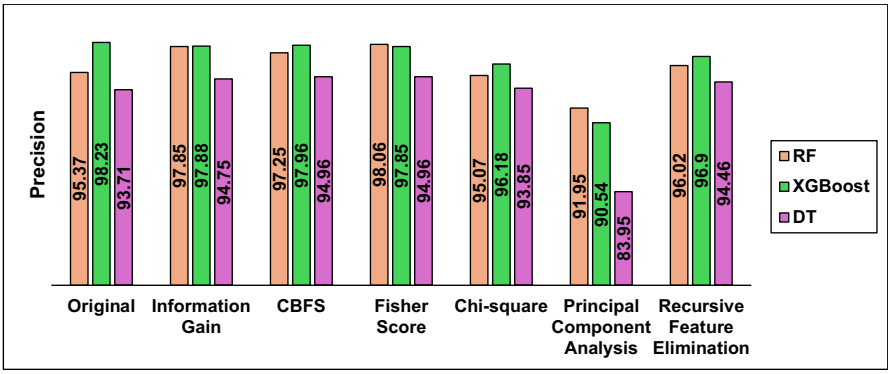


Fig. 3 Figure shows the comparison of precision results of feature selection methods and classification algorithms. X axis is ordered as original, information gain, CBFS, Fisher Score, Chi- Square, Principal Component Analysis, Recursive Feature Elimination. Y axis is Accuracy in percentile. RF, XGBoost, and DT represent Random Forest, eXtreme Gradient Boosting, Decision Tree, respectively

strong and stable performance, especially when combined with Fisher Score and Information Gain, indicating that tree-based ensemble approaches are well-suited for this dataset. In contrast, SVM exhibited substantial performance degradation across most feature selection methods, implying that linear margin-based separation may be insufficient to model the underlying non-linear feature interactions. Although Decision Tree produced relatively consistent results, its performance remained below that of ensemble methods, likely due to its limited capacity to generalize complex decision boundaries.

Figure 3 shows the precision performance of four different classification algorithms using feature selection methods. XGBoost is the best algorithm, generally achieving the highest precision values. Especially with the IG and CBFS methods, XGBoost showed superior performance with precision values of 97,88% and 97,96%, respectively. RF with the FS method showed the highest performance with a precision value of 98,06%. However, SVM generally has lower precision values (shown in Fig. 4). These results show that XGBoost and RF provide the best performance with FS, CBFS, and IG feature selection methods.

In Fig. 4, where the recall performance is presented, it is seen that the XGBoost algorithm achieved high precision values with most feature selection methods. In

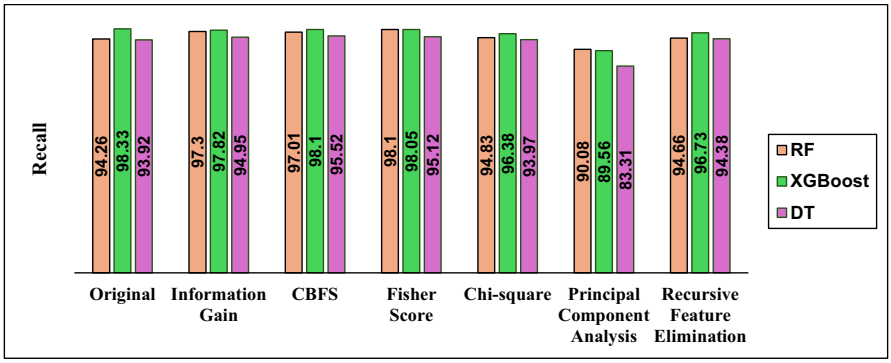


Fig. 4 Figure represents the comparison of recall results of feature selection methods and classification algorithms. X axis is ordered as original, information gain, CBFS, Fisher Score, Chi- Square, Principal Component Analysis, Recursive Feature Elimination. Y axis is Accuracy in percentile. RF, XGBoost, and DT represent Random Forest, eXtreme Gradient Boosting, Decision Tree, respectively

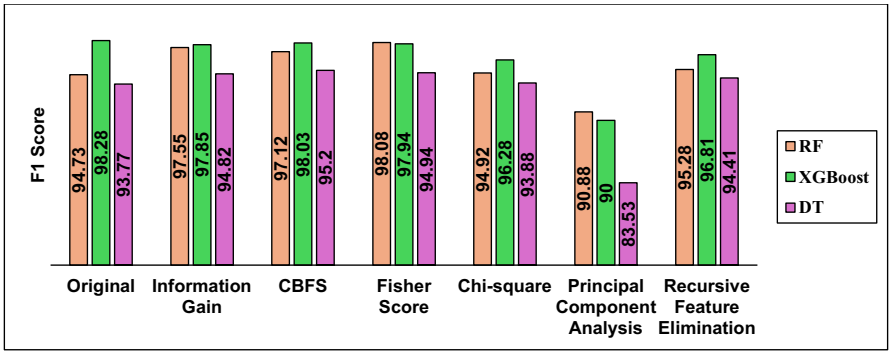


Fig. 5 Figure represents the comparison of F1 score results of feature selection methods and classification algorithms.. X axis is ordered as original, information gain, CBFS, Fisher Score, Chi- Square, Principal Component Analysis, Recursive Feature Elimination. Y axis is Accuracy in percentile. RF, XGBoost, and DT represent Random Forest, eXtreme Gradient Boosting, Decision Tree, respectively

particular, the original dataset (98,33%), CBFS (98,10%), and IG (97,82%) showed the highest performances. RF also showed high sensitivity values with FS (98,10%) and IG (97,30%). SVM has generally low sensitivity values and performed particularly poorly with the CS (54,93%) and FS (55,10%).

In Fig. 5, which shows the F1 score performances, XGBoost achieved the highest F1 scores, especially with the original dataset (98,28%), CBFS (98,03%), and IG (97,85%). RF showed high performance with FS (98,08%) and IG(97,55%). SVM has generally low F1 scores. In conclusion, the best results were obtained with XGBoost-CBFS and RF-FS.

Table 6 Age-restricted analysis of XGBoost performance

Scenario	Model(%)	Accuracy(%)	F1-score(%)
Original	XGB	98.62	98.51
≥65 only	XGB	98.17	97.98
≥65 (No Age Feature)	XGB	96.56	96.38

Table 7 Class-wise performance of the XGBoost model across experimental scenarios

Scenario	Model	Class	F1(%)
Original	XGB	FHm	98,17
Original	XGB	FHs	97,48
Original	XGB	HA	99,31
Original	XGB	NF	99,08
≥65 only	XGB	FHm	97,72
≥65 only	XGB	FHs	97,48
≥65 only	XGB	NF	98,73
≥65 (No Age Feature)	XGB	FHm	98,4
≥65 (No Age Feature)	XGB	FHs	93,71
≥65 (No Age Feature)	XGB	NF	97,04

Robustness and Class-wise Validation of Model Performance

To address the potential confounding effect of age, a sensitivity analysis was conducted by excluding the Healthy Adult group (<65 years) and retraining the model using only participants aged≥65 years. Additionally, the age feature was removed from the feature set. As shown in Table 6, performance remained consistently high (≥65 only: Accuracy 98.17%, F1-macro 97.98%; ≥65 without age: Accuracy 96.56%, F1-macro 96.38%). The limited decrease in performance indicates that model predictions are not solely driven by age, but are substantially supported by postural sway and clinical features.

Class-wise performance analysis(Table 7) demonstrates consistently high precision, recall, and F1-scores across all groups. In the original multi-class setting, all categories achieved F1-scores above 97%, with particularly strong performance for the clinically critical FHm and FHs groups. When the analysis was restricted to participants aged≥65 years, performance remained stable, indicating that discrimination was not driven by the younger HA group. Even after removing the age feature, the model maintained robust classification ability (F1>93% for all classes), with only a modest decrease observed for FHs. Overall, these results confirm that the model reliably distinguishes clinically relevant fall-risk categories beyond simple age-based separation.

Confusion matrices demonstrate that the XGBoost model maintains strong diagonal dominance across all scenarios, with only limited misclassification primarily occurring between the FHm and FHs groups, confirming robust discrimination even after age restriction and removal of the age feature (Fig. 6).

To enhance clinical interpretability, predictive values were calculated for the clinically critical FHm class. In the original dataset, the model achieved a PPV of 96.85% and an NPV of 99.65%, corresponding to approximately 8 false alarms per 1000 screened individuals. In the≥65 restricted cohort, PPV remained high at 96.24% with

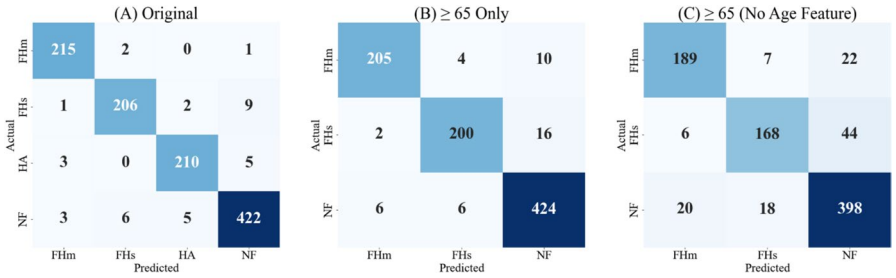


Fig. 6 Figure represents the confusion Matrices of the XGBoost Model Across Experimental Scenarios

an NPV of 97.87%, indicating stable performance even after age-based restriction. These findings suggest that the proposed model maintains low false-positive burden while effectively identifying high-risk individuals, supporting its potential applicability in clinical screening settings.

Discussion

Postural control is known as an essential organization for safely performing daily activities. The loss of balance causes falling—one of the major reasons- for injury resulting in fewer daily activities, decreased functional and social adaptation, and depression (Scheffer et al., 2008). In this study, we consider that postural control which could be affected by aging leads to increased fall risks and therefore to determine the risk of falls in older adults. Since postural control is known as a measurement for evaluating the fall risk, age- related changes in postural sway is related to the capability of maintaining a stable posture. Studies showed a relationship between static body balance measures and risk of falls (Kurz et al., 2013; Melzer et al., 2010) Our results revealed that aging negatively affects postural control, and thereby increases the fall risk. Our finding is in a line with previous studies suggesting that postural sway due to visual impairments (Hytonen et al., 1993) as well as increased imbalance (Kanekar & Aruin, 2014) and fall risks (Santos et al., 2008) in people with older adults Therefore, it is important to understand the multifactorial risk factors helping the improvement evidence-based assessments for older adults to prevent possible falls in the future. In the rehabilittaion purposes, various fall risk tools have been introduced fall risk in different populations including older adults. For example, The John Hopkins Fall Risk Assessment has been developed with intrinsic and extrinsic factors (Poe et al., 2007). Since the most commoly used clinical tools for assessment of fall risk are postural sway and gait parameters (Park, 2018), we analyzed publicly available KINECAL dataset which includes postural sway that could be affected by ageing. Recently, traditional machine learning approaches are capable of handling large datasets but can face challenges (Minvielle, et al., 2017). For example, some studies used ML techniques like tree-based models, logistic regression, RF, SVM, and KNN, and Gboost to identify the significant risk factors in fall prediction models (Lindberg et al., 2020; Liu et al., 2021; Ye et al., 2020a; Young et al., 2021). Since the use of machine learning algorithms to detect health related risks such as fall in

older adults is now usual, the focus of this study was to evaluate the performance of some feature selection methods and classification algorithms using the Kinecal dataset to predict fall risks in older adults. Six feature selection methods were used in the analysis: Information Gain, Correlation Based Feature Selection, Fisher Score, Chi-square, Principal Component Analysis, and Iterative Feature Elimination. These methods were compared with four classification algorithms: Random Forest, Support Vector Machine, XGBoost, and Decision Tree. Metrics such as accuracy, precision, sensitivity, and F1 score were used for performance evaluation.

The results demonstrated that the XGBoost algorithm consistently outperformed other algorithms with most feature selection methods. XGBoost was used to determine the prediction of fall risks using the peoples' demographics and medical records (Ye et al., 2020b) as well as the most crucial gait parameters (Noh et al., 2021). In our study, XGBoost achieved the highest F1 scores with the original dataset (98.28%), CBFS (98,03%), and Information Gain (97,85%). A previous study investigated the standing balance performance by analyzing the trunk sway and postural control test rated by a physical therapist. The outcomes of this study reached %82 classification accuracy (Bao et al., 2019). Furthermore, our findings also demonstrated that the XGBoost algorithm consistently outperformed other algorithms with most feature selection methods.

In our study, Random Forest showed its robustness and ability to handle a variety of datasets, with particularly strong performance with Fisher Score (98,08%) and Information Gain (97,55%). SVM, on the other hand, generally underperformed, especially with the Chi-square and Fisher Score methods, indicating that it is not the best choice for this type of dataset. The Decision Tree gave consistent results, although it underperformed compared to XGBoost and Random Forest. However, it was not able to capture complex interactions within the data as effectively as ensemble methods. We suggest that clinical measurements and sway metrics have been used to estimate fall risk in which our results indicated how machine learning methods can be used in rehabilitation.

In contrast to ensemble-based approaches, SVM demonstrated relatively lower performance across most feature selection methods. This may be attributed to several dataset-specific characteristics. First, the KINECAL dataset includes heterogeneous clinical and postural sway variables that exhibit complex non-linear interactions. Tree-based ensemble methods such as Random Forest and XGBoost are inherently capable of modeling such non-linear relationships without requiring explicit feature transformation, whereas SVM performance is highly dependent on kernel selection and margin optimization. The multi-class structure of the dataset may have further limited SVM performance, as multi-class extensions of SVM rely on pairwise or one-vs-rest strategies that may struggle when class boundaries overlap.

Limitations and Future Directions

Despite the comprehensive comparative framework presented in this study, several limitations should be acknowledged. First, the analysis was conducted using a single publicly available dataset (KINECAL), which may limit the generalizability of the findings. Although stratified k-fold cross-validation provides reliable internal performance estima-

tion, model performance may vary across populations with different demographic characteristics, clinical conditions, and data acquisition protocols. While class imbalance was addressed using SMOTE within the training data, synthetic oversampling may not fully replicate the complexity of real-world minority cases. Although care was taken to prevent information leakage, further validation on larger and independently collected cohorts would strengthen the robustness of the proposed modeling framework. Second, underperformance of SVM algorithm performed poorly compared to other algorithms, particularly with certain feature selection methods like Chi-square and Fisher Score. This underperformance might be due to the KINECAL dataset's heterogeneous clinical and postural sway variables, which exhibit complex non-linear interactions that SVM may struggle to model effectively compared to tree-based ensemble methods. The multi-class structure of the dataset could have further challenged SVM performance. Lastly, this study focused on conventional machine learning algorithms to ensure interpretability and methodological transparency. Deep learning architectures were not explored due to the moderate sample size of the dataset, which may limit the potential benefits of high-capacity models. Future studies are planned to validate the proposed approach using externally collected datasets from diverse clinical environments. Additionally, prospective validation in real-world rehabilitation settings and the integration of advanced modeling techniques may further enhance predictive reliability and translational applicability.

Conclusions

In conclusion, our study comprehensively compared six different feature selection methods (IG, CBFS, FS, PCA, IFE, and Chi-Square) and four classification algorithms (RD, SVM, XGBoost, and DT) to determine the most effective method for the prediction of fall risk. Our results indicate that the XGBoost is highly effective in capturing complex patterns while combining multiple weak learners to transform them into a strong learner in the dataset. The results provide a foundation for future studies focusing on the feature selection methods and classification algorithms for fall risks in different target groups.

Author Contribution All authors contributed equally during the process. Dr. Curuk wrote the main manuscript, M. Pinar was responsible for the data analysis, Dr. E. Pinar was responsible for editing and contributing to the method section, and Dr. E. Demirpolat was responsible for editing and contributing to the discussion section.

Funding Open Access funding provided by the Qatar National Library.

Data Availability The dataset is open source and accessible through a website [25].

Declarations

Ethical Approval Not Applicable.

Disclosures Authors declared no conflict of interest influencing the study reported in this paper.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahmed, Z., et al. (2020) *Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine.* Database-the Journal of Biological Databases and Curation.

Arnold, C. M., et al. (2005). The relationship of intrinsic fall risk factors to a recent history of falling in older women with osteoporosis. *Journal of Orthopaedic & Sports Physical Therapy, 35*(7), 452–460.

Bao, T., et al. (2019). Automatically evaluating balance: A machine learning approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27*(2), 179–186.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST, 25*, 197–227.

Blanchard, R. A., Myers, A. M., & Pearce, N. J. (2007). Reliability, construct validity, and clinical feasibility of the activities-specific fall caution scale for residential living seniors. *Archives of Physical Medicine and Rehabilitation, 88*(6), 732–739.

Bóriková, I. M. T., Miertová, M., & Žiaková, K. (2017). Predictive value of the Morse fall scale. *Central Eur J Nursing Midwifery, 8*(1), 588–595.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Davis, J. C., et al. (2010). International comparison of cost of falls in older adults living in the community: A systematic review. *Osteoporosis International, 21*(8), 1295–1306.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research, 61*, 863–905.

Ferreira, R. N., Ribeiro, N. F., & Santos, C. P. (2022). Fall risk assessment using wearable sensors: A narrative review. *Sensors (Basel), 22*(3), 984.

Freeman, E. E., et al. (2008). Motion-detection threshold and measures of balance in older adults: The SEE Project. *Investigative Ophthalmology & Visual Science, 49*(12), 5257–5263.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics, 28*(2), 337–407.

Fuller, G. F. (2000). Falls in the elderly. *American Family Physician, 61*(7), 2159–2168.

Gárate-Escamila, A. K., El Hassani, A. H., & Andrès, E. (2020). Classification models for heart disease prediction using feature selection and PCA. *Informatics in Medicine Unlocked, 19*, 100330.

Gu, Q., Li, Z., & Han, J. (2012). *Generalized fisher score for feature selection.* arXiv preprint arXiv: 1202.3725.

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

Hall, M. A. (1999). *Correlation-based feature selection for machine learning* in *The University of Waikato*.

Hendrich, A. L., Bufalino, A., & Groves, C. (2020). Validation of the Hendrich II Fall Risk Model: The imperative to reduce modifiable risk factors. *Applied Nursing Research, 53*, 151243.

Horak, F. B., Nashner, L. M., & Diener, H. C. (1990). Postural strategies associated with somatosensory and vestibular loss. *Experimental Brain Research, 82*(1), 167–177.

Howcroft, J., Kofman, J., & Lemaire, E. D. (2013). Review of fall risk assessment in geriatric populations using inertial sensors. *Journal of Neuroengineering and Rehabilitation, 10*, 91.

Hytonen, M., et al. (1993). Postural control and age. *Acta Oto-Laryngologica, 113*(2), 119–122.

Jadhav, S., He, H., & Jenkins, K. (2018). Information gain directed genetic algorithm wrapper feature selection for credit rating. *Applied Soft Computing, 69*, 541–553.

Jahn, K., Zwergal, A., & Schniepp, R. (2010). Gait disturbances in old age: classification, diagnosis, and treatment from a neurological perspective. *Deutsches Ärzteblatt International, 107*(17), 306–15.

Kanekar, N., & Aruin, A. S. (2014). The effect of aging on anticipatory postural control. *Experimental Brain Research, 232*(4), 1127–1136.
Kim, J. W., et al. (2010). Sex differences in the postural sway characteristics of young and elderly subjects during quiet natural standing. *Geriatrics & Gerontology International, 10*(2), 191–198.
Kurz, I., Oddsson, L., & Melzer, I. (2013). Characteristics of balance control in older persons who fall with injury - A prospective study. *Journal of Electromyography and Kinesiology, 23*(4), 814–819.
Lindberg, D. S., et al. (2020). 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. *International Journal of Medical Informatics, 143*, 104272.
Liu, C. H., Hu, Y. H., & Lin, Y. H. (2021). A machine learning-based fall risk assessment model for inpatients. *Computers in Informatics Nursing, 39*(8), 450–459.
Marigold, D. S., et al. (2004). Contribution of muscle strength and integration of afferent input to postural instability in persons with stroke. *Neurorehabilitation and Neural Repair, 18*(4), 222–229.
Maudsley-Barton, S., & Yap, M. H. (2023). KINECAL: A dataset for falls-risk assessment and balance impairment analysis. *Scientific Data, 10*(1), 633.
Melzer, I., Kurz, I., & Oddsson, L. I. E. (2010). A retrospective analysis of balance control parameters in elderly fallers and non-fallers. *Clinical Biomechanics, 25*(10), 984–988.
Millet, A., et al. (2023). Machine learning techniques applied to the development of a fall risk index for older adults. *IEEE Access, 11*, 84795–84809.
Minvielle, L., et al. (2017). *Fall detection using smart floor sensor and supervised learning.* 2017 39th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), 2017: p. 3445–3448.
Myers, H. (2003). Hospital fall risk assessment tools: A critique of the literature. *International Journal of Nursing Practice, 9*(4), 223–235.
Noh, B., et al. (2021). XGBoost based machine learning approach to predict the risk of fall in older adults using gait outcomes. *Scientific Reports, 11*(1), 12183.
Oliver, D., et al. (1997). Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: Case-control and cohort studies. *BMJ, 315*(7115), 1049–1053.
Organization, W.H. *tep Safely: Strategies for Preventing and Managing Falls Across the Life-Course*. 2021 22.02.2026]; Available from: https://iris.who.int/server/api/core/bitstreams/e7aabc9b-6978-43eb-a6b7-5172d11ce379/content
Park, S. H. (2018). Tools for assessing fall risk in the elderly: A systematic review and meta-analysis. *Aging Clinical and Experimental Research, 30*(1), 1–16.
Peterka, R. J. (2002). Sensorimotor integration in human postural control. *Journal of Neurophysiology, 88*(3), 1097–1118.
Pinar, M., Aktas, A., & Ulku, E. E. (2025). Feature efficiency in IoMT security: A comprehensive framework for threat detection with DNN and ML. *Computers in Biology and Medicine, 186*, 109603.
Poe, S. S., et al. (2007). The Johns Hopkins Fall Risk Assessment Tool: Postimplementation evaluation. *Journal of Nursing Care Quality, 22*(4), 293–298.
Qiu, H., et al. (2018). Application of wearable inertial sensors and a new test battery for distinguishing retrospective fallers from non-fallers among community-dwelling older people. *Science and Reports, 8*(1), 16349.
Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine, 47*(1), 31–39.
Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*(19), 2507–2517.
Santos, B. R., et al. (2008). Reliability of centre of pressure summary measures of postural steadiness in healthy young adults. *Gait & Posture, 27*(3), 408–415.
Scheffer, A. C., et al. (2008). Fear of falling: Measurement strategy, prevalence, risk factors and consequences among older persons. *Age and Ageing, 37*(1), 19–24.
Schmid, N. A. (1989). Reducing patient falls: A research-based comprehensive fall prevention program. *Military Medicine, 155*, 5.
Shuja, M., Mittal, S., & Zaman, M. (2019). Effective prediction of type ii diabetes mellitus using data mining classifiers and SMOTE. *Advances in Computing and Intelligent Systems, Icacm, 2020*, 195–211.
Silva, J., et al. (2017). *Comparing machine learning approaches for fall risk assessment.* Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies, Vol 4: Biosignals, p. 223–230.

Song, F., Guo, Z., & Mei, D. (2010). *Feature selection using principal component analysis.* , in *In 2010 international conference on system science, engineering design and manufacturing informatization*. 2010, IEEE. p. 27–30.

Torres, C. H., & Hsu, T. (2017). Comprehensive geriatric assessment in the older adult with cancer: A review. *European Urology Focus, 3*(4–5), 330–339.

Wang, M., & Chen, H. (2020). Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Applied Soft Computing, 4*(3), 211–229.

Ye, C. Y., et al. (2020a). Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. *International Journal of Medical Informatics*. https://doi.org/10.1016/j.ijmedinf.2020.104105

Ye, C., et al. (2020b). Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. *International Journal of Medical Informatics, 137*, 104105.

Young, A. J., et al. (2021). Using machine learning to make predictions in patients who fall. *Journal of Surgical Research, 257*, 118–127.

Zhai, Y., Song, W., Liu, X., Liu, L., & Zhao, X. (2018). *A chi-square statistics based feature selection method in text classification*, in *In 2018 IEEE 9th International conference on software engineering and service science (ICSESS)*. p. 160–163.

Zhang, L., et al. (2015). *Classification of older adults with/without a fall history using machine learning methods.* 2015 37th Annual International Conference of the Ieee Engineering in Medicine and Biology Society (Embc), p. 6760–6763.

Zhou, J., Gandomi, A. H., Chen, F., & Holzinger, A. (2021). Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics, 10*(5), 593.

Zurbuchen, N., Wilde, A., & Bruegger, P. (2021). A machine learning multi-class approach for fall detection systems based on wearable sensors with a study on sampling rates selection. *Sensors (Basel), 21*(3), 938.

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

Authors and Affiliations

Etem Curuk[1] · Merve Pinar[2] · Ertugrul Pinar[3] · Eyyup Demirpolat[4]

✉ Etem Curuk
ecuruk@qu.edu.qa

✉ Merve Pinar
merve.pinar@marmara.edu.tr

Ertugrul Pinar
ertugrulpinar@gmail.com

Eyyup Demirpolat
edemirpolat@bingol.edu.tr

[1] The Department of Rehabilitation Sciences, College of Applied Health Sciences, Qatar University, Doha, Qatar

[2] Department of Computer Engineering, Faculty of Technology, Marmara University, Istanbul, Turkey

[3] Neurosurgery Clinic, Hisar Hospital Intercontinental, Istanbul, Turkey

[4] Department of Therapy and Rehabilitation, Vocational School of Health Services, Physiotherapy Program, Bingöl University, Bingöl, Turkey