Exploratory Analysis of Machine Learning Methods for the Prognosis of Falls in Elderly Care Based on Accelerometer Data

Lukas Klein[1,2(✉)], Christoph Ostrau[1], Michael Thies[2], Wolfram Schenck[1], and Ulrich Rückert[2]

[1] Center for Health, Social Affairs and Technology (CareTech OWL), University of Applied Sciences and Arts, Bielefeld, Germany
{lukas.klein,christoph.ostrau}@hsbi.de
[2] Center for Cognitive Interaction Technology (CITEC), Bielefeld University, Bielefeld, Germany

Abstract. This paper investigates the feasibility of employing machine learning techniques to categorize individuals into fall-risk and non-fall-risk groups based solely on accelerometer data. The research utilizes a publicly available movement monitoring dataset, containing accelerometer data from a diverse group of individuals. The study pursues three primary objectives. First, it develops a preprocessing pipeline to prepare raw accelerometer data, which includes noise reduction, data cleaning, and identification of walking segments and the extraction of over twenty gait-related features. The second objective is to systematically explore the influence of these features on machine learning model performance. Gait stability-related parameters, known from medical literature, are of particular interest. To fulfil this objective, different machine learning algorithms are evaluated using an automated exploration framework. The third objective centres on finding a balanced combination of features and lightweight machine learning models suitable for embedded systems, which typically have limited computational resources. The emphasis here is on computational efficiency, an original aspect of this study. The results indicate that gradient boosting algorithms, such as XGBoost, LightGBM, and CatBoost, outperform other models, achieving promising performance results, including an area under the curve (AUC) score of up to 0.949.

Keywords: Machine Learning · Optimization · AutoML · Gait Analysis · Fall Risk Assessment · Feature Engineering

1 Introduction

Falls, especially among older individuals, are a serious problem in an ageing society. Elderly people are more vulnerable to falls due to various factors [21], leading to significant personal consequences such as reduced quality of life, loss

D. Salvi et al. (Eds.): PH 2023, LNICST 572, pp. 423–437, 2024.
https://doi.org/10.1007/978-3-031-59717-6_27

of autonomy, reduced social participation, chronic pain, and even hospitalization or mortality [22]. According to the Centers for Disease Control and Prevention (CDC), a staggering 27.5% of adults aged 65 and older in the United States experienced a fall in 2018, with around 24% of these incidents resulting in fall-related injuries [13]. Falls rank as the primary cause of injuries among adults aged 65 and older in the United States, accounting for approximately 3 million emergency department visits and 32,000 fall-related deaths annually. As the number of falls rises, so do the associated injuries and, consequently, healthcare costs [7]. To mitigate the personal, economic, and societal consequences, reducing falls among this vulnerable age group is imperative. The most effective approach to reducing falls is an early identification of individuals at risk, enabling preventive actions. However, fall risk assessments are typically conducted retrospectively through periodic medical examinations involving comprehensive questionnaires and laboratory tests [17]. To streamline ubiquitous fall risk assessment, modern sensor technology and advanced *machine learning (ML)* algorithms can be utilized (e.g. [6]). Promising approaches rely on three-dimensional motion data recorded by accelerometers, using ML techniques to predict fall risks accurately [1,24]. If these models can reliably identify fall risk, they could be integrated into cost-effective embedded systems which are continuously wearable by individuals, akin to small smartwatches. Such tools could serve as early warning systems, alerting wearers to changes in their movement patterns indicative of an increased risk of falling. This opens up the possibility of timely medical interventions to prevent impending falls and their associated consequences.

This paper assesses the applicability of ML methods for categorizing patients into two groups: those at risk of falling and those not at risk, based solely on accelerometer data from the publicly available *Long Term Movement Monitoring Database v1.0.0 (LTMM)* dataset [25]. To achieve this, we establish a comprehensive ML pipeline. The initial step is to create an effective preprocessing pipeline. This pipeline prepares raw accelerometer data by reducing noise, cleaning sensor data, and extracting walking segments. Next, we calculate features related to gait characteristics and examine their influence on ML model performance. We systematically explore various features and ML algorithms to identify an optimized combination, with a focus on lightweight models suited for resource-constrained embedded systems. In doing so, our scope does not involve implementing an embedded system, but rather evaluating the methods' feasibility and assessing optimal features in terms of computational requirements. This focus is a crucial part of our research objectives and a novel contribution to this specific application context. Furthermore, our study seeks to explore various ML techniques and to exhibit their ability to differentiate between fallers and non-fallers, based on data obtained through a single accelerometer and a constrained dataset, without extensively optimizing model architecture or preprocessing steps.

The remainder of this paper is structured as follows: In Sect. 2, the related work and research is presented. Then, the methods used to set up a preprocessing pipeline and to explore ML algorithms are explained in Sect. 3. The main findings are presented and discussed in Sect. 4 and Sect. 5 respectively. The paper is concluded in Sect. 6.

2 Related Work

Several papers have been published on the study of the LTMM dataset, as well as on predicting the risk of falling in individuals using ML algorithms with accelerometer data. The Weiss et al. research team [25], responsible for acquiring the LTMM dataset, analysed whether the acceleration data and its respective parameters display statistical correlation to the subjects' fall status in their initial publication on the dataset. They determined whether the accelerometer data of the dataset and the parameters derived from these can be used to assess the risk of falling in general. Weiss et al. calculated so-called gait-specific parameters [2] from walking segments of the raw sensor data, and performed statistical tests to compare these parameters between the class of fallers and the control group. They then used simple logistic regression to investigate the ability of the different parameters to identify the fall risk of the subjects. Weiss et al. found that there were statistically significant correlations between the sensor data and the calculated gait parameters and the subjects' fall risk status. Another paper on this dataset, published by the research team led by Ihlen et al. examined a new measure of gait stability in terms of its ability to discriminate between the fallers and the control group. The so-called'phase-dependent local dynamic stability' (λ) [10] measures how a subject's gait responds to infinitesimally small perturbations. Ihlen et al. state that this measure is very good at discriminating between fallers and non-fallers. In particular, the phase-dependent λ between 0% and 60% of the gait cycle significantly improved discrimination performance. In combination with 38 more conventional gait parameters used in the aforementioned Weiss et al. paper [25], Ihlen et al. achieved an *area under the curve (AUC)* score of 0.93 using partial least squares discriminant analysis. This is a very high value and close to a perfect result, and can be used as a reference when exploring other ML models. Van Schooten et al. [19] obtained a much larger dataset of 169 patients. They collected sensor data from an accelerometer worn on the lower back for 7 consecutive days and examined the predictive ability of features calculated from the accelerometer sensor. Using logistic regression, they obtained an area under the curve of 0.82. This result was obtained by combining data from the sensor and questionnaire data from the subjects. Other ML methods have been applied in the literature by other research groups on different data sets. Howcroft et al. [9] trained a motion dataset for fall risk prediction in older adults using multiple accelerometers at different locations on the human body. Howcroft et al. achieved an accuracy detecting people that have a higher risk of falling based on their fall history of 57%, a sensitivity of 35% and a specificity of 67%. They claimed that a fall risk screening tool should use multi-sensor data, as combining the data from all sensors improved the sensitivity of the best performing model (a neural network) to 43% and accuracy to 57%, while specificity dropped slightly to 65%. However, if one is limited to using only one sensor, they suggested attaching that sensor to the pelvic location, which is approximately the same location as the sensor worn in the LTMM data collection process. The research group of Aicha et al. [1] investigated deep learning methods to predict falls in older adults, also using accelerometer data from the lower trunk of the human body. They reached

a peak performance of an AUC of 0.75 with intensive preprocessing of the data. A recent study investigated fall prediction in patients with Parkinson's disease [24], utilizing real-world data collected from foot-worn inertial sensors. Patients undertook several unsupervised 10-m walking tests daily, and data was collected over a two-week period. Patients were required to self-report any severe falls over the course of three months following the data collection period. Employing a random forest algorithm, the authors achieved a sensitivity (recall) rate of 60% and specificity rate of 88%, resulting in a balanced accuracy rate of 74%.

3 Methodology

In this chapter, we introduce the Long Term Movement Monitoring Database dataset, which comprises three-dimensional acceleration data collected from daily activities of older adults. Additionally, we provide an overview of the essential preprocessing pipeline and the process of feature engineering, setting the stage for the subsequent exploration of ML methods for fall risk prediction. Moreover, MLJAR [16], an automated ML library, is employed to delve into multiple ML algorithms and perform feature engineering.

3.1 Dataset

Traditionally, assessing the risk of falls in the elderly relied on subjective self-reports or isolated assessments, lacking objectivity. In response, the LTMM dataset [25], collected in 2016 by a consortium of researchers from various institutions, aimed to explore whether 3-dimensional acceleration data from everyday life could provide insights into the fall risk of older adults. This dataset is publicly accessible on PhysioNet [8], managed by MIT's Laboratory for Computational Physiology and supported by the National Institute of Biomedical Imaging and Bioengineering, allowing for sharing, modification, and use under the Open Data Commons Attribution Licence v1.0. The data collection process involved equipping 71 older adults, aged 65 to 87, with DynaPort Hybrid sensors from the Dutch company McRoberts. These sensors were affixed to the lower back, specifically at the fifth lumbar vertebra, for a continuous three-day monitoring period. The sensors recorded three-axis acceleration data (vertical, mediolateral, and anterior-posterior) via accelerometers and yaw, pitch, and roll velocities through triaxial gyroscopes, all sampled at a rate of 100 Hz. Based on the approaches in the literature [25] in order to make both the preprocessing pipeline and the learning algorithms least computationally intensive, only the the-axis acceleration data of the dataset was used. None of the participants had been diagnosed with balance or cognitive impairments. They were categorized as fallers or non-fallers based on self-reported fall history, with fallers having experienced two or more falls in the previous year. There were no noteworthy differences in demographic factors, such as age, gender, social status, height, weight, or body mass index, between the two groups of fallers and non-fallers. The data collection process encompassed four phases:

1. traditional laboratory-based fall risk assessments, including tests like the Dynamic Gait Index (DGI), Berg Balance Scale (BBS), Timed Up and Go test (TUG), Four Square Step Test (FSST), Mini Mental State Examination (MMSE), and Activity-specific Balance Confidence scale (ABC)
2. a laboratory gait assessment where participants walked for a minute with the same sensor belt worn during the three-day monitoring
3. three days of sensor wear in their daily routines, allowing for sensor removal during specific activities
4. a follow-up period of six months during which participants reported any falls, aiming to assess the predictive potential of accelerometer data.

While Weiss and colleagues [25] reported no significant differences in walking duration between the two participant groups, minor variations in sensor wear time, including weekdays versus weekends, can introduce variations in signal preprocessing and filtering of walking segments. These discrepancies may result in an uneven distribution of walking segments within the database, impacting the training of ML algorithms. Consequently, it is crucial to consider these potential imbalances when evaluating ML methods using diverse evaluation metrics. The data set was split into test data and training data using an 80/20 split. No explicit care was taken to perform the split according to the subjects. However, since the data was not shuffled before the split, it can be considered that the test and training data still is split according to the subjects, except for the data of the subject which may fall into the split boundary.

3.2 Preprocessing Pipeline

To employ ML methods for classifying sensor data into fallers and non-fallers, preprocessing of the raw sensor data is essential. This preprocessing, inspired by the work of Ullrich et al. [23], encompasses the three stages of signal preprocessing, movement detection, and frequency analysis. The aim of this process is to identify all data segments where the test subject is walking for a minimum of one minute. Signal preprocessing involves the removal of outliers, noise, and gravity components through high- and low-pass filters to ensure data integrity. The full preprocessing pipeline is depicted in Fig. 1. The preprocessed data undergoes segmentation using a sliding window technique with a 60-s window length and 50% overlap, enhancing temporal resolution, which is motivated by related work [18,26]. Movement detection entails determining if a data window contains movement of any kind, employing the Signal Magnitude Area threshold method as a pre-filtering step to exclude non-active sections. The *Signal Magnitude Area (SMA)* quantifies the intensity of a time-varying signal, commonly applied in physical activity analysis [12]. For a dataset with n values, it is computed as the sum of the absolute amplitudes of the signal within a specified time window. In the context of physical activity analysis, SMA is utilized to measure movement intensity and energy expenditure. For continuous time sensor data in the LTMM dataset, SMA computation is adapted as follows:

$$\mathrm{SMA} = \frac{1}{T} \int_0^T \left[|X(t) - \mu_x| + |Y(t) - \mu_y| + |Z(t) - \mu_z| \right] dt \ , \tag{1}$$

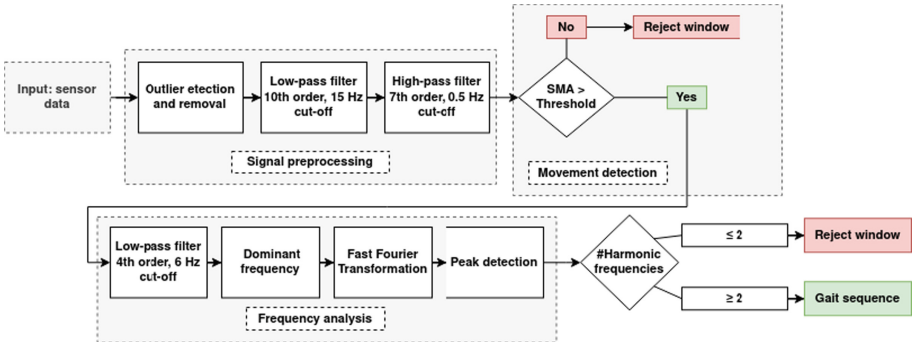

Fig. 1. Algorithm flowchart of the proposed data preprocessing and walking segment detection.

where T denotes the length of the sliding window filter, X, Y, Z represent the linear acceleration in the vertical (X), mediolateral (Y) and anterior-posterior (Z) axes. μ_i denotes the average linear acceleration of the respective axis and t represents the time step.

SMA values are evaluated against a threshold to identify movement within a given time window. Various methods are explored to establish this threshold, including visual observation, predefined thresholds from prior studies, averaging SMA values from laboratory data, and individual patient-specific thresholds. The aim is to find a threshold strategy that balances simplicity and computational efficiency while effectively detecting any activity. Subsequently, frequency analysis is performed on time windows that passed the movement detection stage. A fourth-order low-pass filter (cut-off at 6 Hz) is applied to isolate harmonic frequencies indicative of walking. The dominant frequency is computed through autocorrelation and analysed using the discrete Fourier Transform. Windows are classified as walking or non-walking based on the presence of harmonic frequencies. It is worth noting that frequency analysis is conducted solely along the vertical axis of motion, a pragmatic choice balancing accuracy and computational efficiency.

3.3 Feature Engineering

Feature engineering plays a pivotal role in ML workflows by transforming raw data into informative variables or features that enhance model performance. This process encompasses techniques such as feature extraction, selection, scaling, and transformation to represent data optimally for ML algorithms. It is iterative, relying on domain knowledge and exploration, and aims to minimize feature collinearity, maximize class separability, and reduce computational complexity. In this study, an extensive feature engineering pipeline is implemented, particularly emphasizing feature extraction. Features include gait parameters such as stride regularity, step characteristics, dominant frequencies, sensor data statis-

tics, Signal Magnitude Area, mean swing time, time between peaks, displacement, and *Local Dynamic Stability (LDS)*. All calculated features are depicted in Table 1. These features are prepared for training with proper scaling, ensuring their effectiveness in the ML models.

To investigate the impact of the calculated feature on the performance of the ML methods, two different feature importance metrics are analysed. Firstly the permutation-based feature importance, a model inspection technique that assesses a model's reliance on specific features by measuring the reduction in its performance when individual feature values are randomly shuffled [4]. Secondly, the SHAP [11] importance scores that quantify the contribution of each feature to a model's prediction by considering all possible feature combinations and their impact on predictions based on the Shapley values [20].

Table 1. Overview of the feature that are calculated from the sensor data and used for learning with ML methods.

Feature	Abbreviation	Comment
Local Dynamic Stability	lambda_diff	Local Dynamic Stability of time window
Range of Sensor Data	acc_range_{v,ml,ap}	Range of sensor data values in each axis
Average of Sensor Data	acc_avg_{v,ml,ap}	Average of sensor data in each axis
SMA Value	SMA_value	SMA value of time window
Dominant Frequencies	{v,ml,ap}_dom_freq	Dominant frequencies of time window of each axis
Time between Peaks	time_between_peaks	Time interval between two consecutive peaks in accelerometer signal
Step Time Variation Coefficient	step_time_var_coeff	Ratio of standard deviation and mean of time between steps
Mean Step Time	mean_step_time	Average time of each step in a time window
Step Time Variation	step_tim_var_sd	Standard deviation of the step times
Number of Steps	num_steps	Number of steps in a time window
xCoM Displacement	xcom_{v,ml,ap}_displacement	Change in position of the extrapolated center of mass of the subjects
Cadence	cadence	Number of steps in a specific time period
Step Symmetry	step_symmetry	Quantifies the similarity between the movement of both feet
Mean Swing Time	mean_swing_time	Average duration of the swing phase of a step

3.4 Validation of Data Preprocessing

Validating the accuracy of identified walking segments is crucial for subsequent analyses. It ensures that the features derived from these segments are robust and contribute effectively to ML models. Validation methods include visual inspection, comparison with similar studies (limited by data availability), and an inductive approach where classical gait parameters are calculated and compared with expected values, providing confidence in the authenticity of the segments.

3.5 Exploration of ML Methods

MLJAR [16] is an *automated ML (AutoML)* library that streamlines the entire ML workflow. It is designed to facilitate intensive data analysis and comparative evaluation of various ML methods. In this study, MLJAR is leveraged to explore several ML algorithms, including Decision Trees, Random Forests, Extra

Trees, XGBoost, LightGBM, CatBoost, and simple Neural Networks consisting of fully connected layers. For training sessions, the dataset is partitioned into training and test sets using an 80/20 split. The test set is reserved for final model evaluation exclusively. In addition, a 5-fold cross-validation on the training data is conducted to assess each algorithm's performance and feature importance. MLJAR also offers preprocessing techniques like generating'Golden Features' and conducting'Feature Selection' to further enhance model performance. For the former, MLJAR generates unique feature pairs from the original input features and combines them with subtraction or division operators to obtain new features. For each (generated) feature, an importance score is calculated, the features are ranked according to their importance score, and the most important features are implanted into the training data. For Feature Selection, MLJAR inserts a random feature into the training data and trains the yet best model with this random features included. For each original feature, MLJAR calculates how many times its importance on the performance is smaller than the importance of the random feature. Every feature, that is at least on more than half the learners less important than the randomly generated feature, gets dropped from further learning.

4 Results

In this section, we delve into the outcomes of our sensor data preprocessing pipeline. Additionally, we explore feature engineering, emphasizing the significance of different features, and subsequently, we evaluate various ML models for binary classification, considering diverse evaluation metrics and the influence of feature engineering on model performance.

4.1 Walking Segment Detection

In this section, we present and analyse the results of our sensor data preprocessing pipeline, particularly focusing on the critical parameter of the Signal Magnitude Area (SMA) threshold. We experimented with various SMA thresholds, ranging from 0 to 1.0, to find the optimal value for the initial stage of preprocessing. Visual inspection of resulting movement segments was used for assessment. Using individual SMA thresholds based on each patient's 60-s laboratory walking segment data did not yield satisfactory results, as well as utilizing the mean SMA value from all participants' laboratory data, since it worsened the imbalance of the resulting dataset. The best outcome was achieved with an SMA threshold of 0.2, striking a balance between filtering stringency and dataset balance. Using this as a threshold for SMA-based movement detection, the full preprocessing pipeline resulted in an imbalanced dataset consisting of 5,951 walking segments. Of these, 68% were control group samples, whilst 32% were faller group samples. We validated the preprocessing by deriving classical gait parameters from the identified walking segments, specifically the average step duration and the dominant frequencies in the three axes of motion. The

identified walking segments have an average step duration of 0.5 s, placing it in the lower range of average step duration for adults, which is between 0.49 and 0.59 s. [3,14,15]. The average dominant frequency in the vertical axis, the posterior-anterior axis and the mediolateral axis of the identified walking segments is with an average of 2.5 Hz, 2.3 Hz respectively 1.8 Hz within the 1–3 Hz which is reported as the normal range of dominant frequencies for adults [5]. The majority of values for the identified walking segments fall within the range of values found in literature, resulting in a false discovery rate of approximately 3%. However, the literature values were reported for adults of all ages, while our dataset includes solely older individuals who may have a higher risk of falling. This suggests that our preprocessing pipeline effectively extracts walking segments from sensor data.

4.2 Feature Engineering

In evaluating feature importance for predicting fall risk, we primarily examine the output of the best-performing models in MLJAR. The goal is to identify which features play a crucial role in the prediction task. Based on permutation-based importance plots, depicted in Fig. 2, the feature lambda_diff, representing the mean local dynamic stability of walking segments, consistently stands out as the most important, with weights ranging from approximately 0.16 to 0.22 across different learners. The range of vertical acceleration values follows as the second most important feature, with weights around 0.10. The average swing time feature, valued at approximately 0.10, also proves significant in both CatBoost and XGBoost models. Additionally, we consider SHAP importance scores, which account for feature interactions. These scores reaffirm the dominance of lambda_diff as the most critical feature, with an average weight of about 1.1 across all learners. The range of vertical acceleration values remains highly relevant. Notably, the feature cadence and the average accelerations in mediolateral and anterior-posterior directions hold very low importance in the models. This is likely because cadence is closely related to the number of steps, introducing redundancy. Ablation studies demonstrate that models lacking the three most important features (lambda_diff, vertical acceleration range, and mean swing time) perform significantly worse. This underscores the significance of these aforementioned features. Conversely, models solely trained on these top three features exhibit reduced effectiveness. While MLJAR offers automated feature generation techniques (compare Sect. 3.5), models that perform best tend to rely on manually engineered features. Nonetheless, models utilizing automatically generated features still achieve reasonable performance, suggesting their relevance. Among the generated features, those created using the Golden Features technique are the most impactful, particularly when they combine features with high importance scores. The difference between lambda_diff and time_between_peaks stands out as the most significant generated feature, reinforcing the significance of lambda_diff. In summary, lambda_diff, vertical acceleration range, and mean swing time are key features in predicting fall risk, while other features, such as

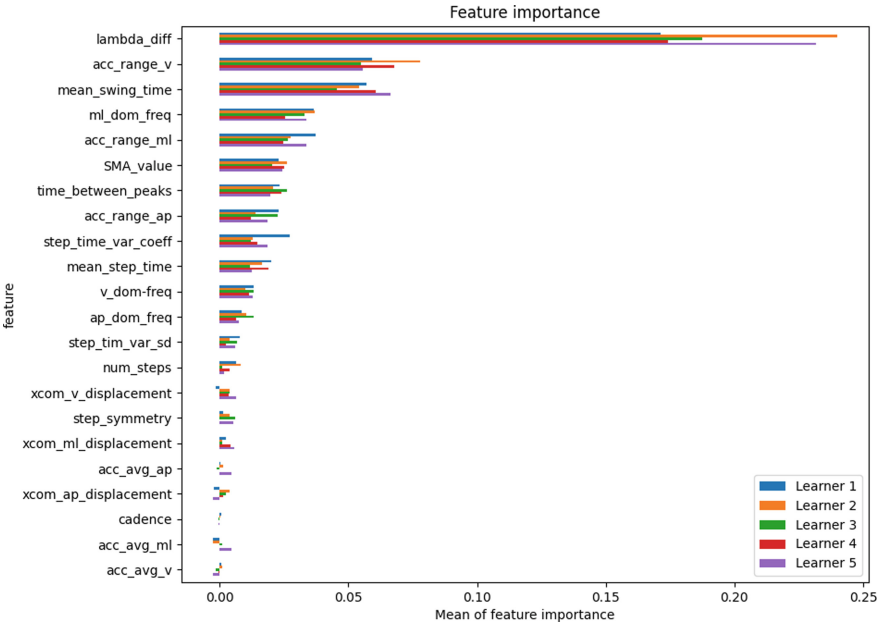

Fig. 2. Permutation based feature importance plots of best performing XGBoost model. Refer to Table 1 for an explanation of the features.

cadence, hold little importance. Automated feature generation, while less frequently used in the best models, still contributes to reasonable performance, especially when combining high-importance features. However, to achieve optimal model performance, it is advisable to utilize all manually engineered features, which is applied in the remainder of this paper.

4.3 Evaluation of ML Models

This section presents the outcomes of training runs conducted using MLJAR and identifies the most suitable ML techniques for addressing the binary classification task at hand. To determine the optimal approach for our dataset, we leveraged the insights gained from feature engineering evaluation and preprocessing. Instead of using raw sensor data, we utilized feature vectors extracted from walking segments as input data. The choice of evaluation metric significantly affects the model performance results. While log loss emerged as the top metric to optimize for to get the best results overall, we considered additional metrics like AUC, F1-score, recall, *Matthews Correlation Coefficient (MCC)* and precision due to the dataset's slight imbalance and the critical nature of the problem, where the identification of individuals at high risk of falling is crucial. To investigate the impact of using engineered features versus raw sensor data, we initially trained models solely on raw sensor data values, leading to an enormous dataset size. This approach significantly increased training times, for instance, taking

more than 17 min to train a CatBoost model, compared to just 17 s when using feature vectors. Furthermore, the best results of MLJAR on the raw sensor data of walking segments also yield significantly worse results with an AUC of 0.7358, a MCC score of 0.252 and a very low recall score of 0.255 (CatBoost).

Consequently, we observed that feature engineering plays a pivotal role in model performance. The results indicated that tree-based algorithms, specifically Extra Trees and Random Forests, underperformed compared to gradient boosting methods such as LightGBM, XGBoost, and CatBoost across various evaluation metrics. Furthermore, these tree-based models exhibited longer single prediction and training times, making them less suitable for the task.

Table 2. Comparison of evaluation metrics of the best scores of CatBoost (depth: 8, random subspace method value: 0.9–1), LightGBM (number of leaves: 100–127, minimal data in leave: 10–20), XGBoost (maximal depth: 8, minimum child weights: 5) and Neural Network models (fully-connected feed-forward network, size: $22 \times 16 \times 32 \times 1$). Each cell contains the metric value.

Metric	CatBoost	LightGBM	XGBoost	Neural Network
Log Loss	**0.264**	0.290	0.291	0.415
AUC	**0.949**	0.939	0.937	0.905
F1	**0.800**	0.797	0.780	0.703
Accuracy	**0.884**	0.877	0.873	0.842
Precision	**0.858**	0.816	0.842	0.837
Recall	0.750	**0.780**	0.726	0.607
MCC	**0.723**	0.710	0.695	0.614
Predict time (s)	**0.028**	0.060	0.056	0.028
Train time (s)	81.16	28.30	31.24	**9.58**

Neural Networks, while showing reasonable performance, fell short of the best gradient boosting models in terms of metrics like log loss, F1-score, and AUC. Additionally, they incurred significantly longer prediction times, making them less efficient. The three gradient boosting techniques, LightGBM, XGBoost, and CatBoost, and a Neural Network were further examined to optimize their performance using MLJAR's 'Optuna' and 'Perform' modes. These results are displayed in Table 2. CatBoost outperformed its counterparts in log loss and AUC, achieving superior scores. LightGBM, while performing well, came in second place in these metrics. CatBoost also exhibited the highest precision, while LightGBM excelled in recall. To address class imbalance, precision-recall curves were analysed, with CatBoost demonstrating the best performance. All models displayed normal precision-recall curves without anomalies. Regarding memory consumption, CatBoost models were found to require larger file storage sizes on average, suggesting higher memory requirements compared to XGBoost and LightGBM models. In summary, gradient boosting algorithms, particularly CatBoost, LightGBM, and XGBoost, exhibited superior performance in classifying

individuals at risk of falling compared to other methods. They achieved excellent results with very low prediction times, making them suitable for real-time applications. Neural Networks, while capable of achieving reasonably good results, lagged behind gradient boosting methods. Tree-based algorithms, such as Extra Trees and Random Forests, were less efficient in terms of both performance and computational speed.

5 Discussion

The data preprocessing pipeline implementation detected 5,951 walking segments, each of which were 60 s long. 68% of these segments belonged to the control group, while 32% were associated with a higher risk of falling. This preprocessing, including movement detection and feature extraction, significantly improved the accuracy of ML models for classifying individuals as fallers or controls. To enhance data quality, we applied high-pass and low-pass filters to the sensor data, reducing noise and increasing the signal-to-noise ratio. Additionally, outlier detection was employed to remove anomalous data points (e.g. caused by readout errors). However, the impact of these filtering techniques on ML model performance remains to be thoroughly investigated. These preprocessing steps allowed us to extract meaningful gait parameters from walking segments, which were crucial for fall risk assessment. We compared models trained on these engineered features with models using raw sensor data. The former consistently outperformed the latter, emphasizing the importance of feature engineering, particularly for tree-based ML algorithms. While we have diligently computed all relevant features and tested their combinations to the best of our knowledge, we must acknowledge the potential for further features that might improve ML performance. Additionally, our research focused solely on the LTMM dataset, which presents a comparatively small sample of individuals wearing the sensors. Future research could explore the impact of a more diverse dataset with a larger pool of test subjects, such as the dataset obtained by van Schooten et al., which provides data over a longer duration [19].

Another key finding is the identification of essential features for fall risk assessment, such as mean local dynamic stability, vertical acceleration range, and average swing time. Interestingly, cadence showed little significance in model performance. These results are consistent with previous studies and demonstrate the importance of stability-related features. Our analysis of various ML algorithms revealed that gradient boosting frameworks (XGBoost, CatBoost, and LightGBM) consistently achieved the best results, with slight variations in evaluation metrics. Table 2 displays the most effective models, presenting AUC scores of up to 0.949 (CatBoost). These findings surpass the outcomes of previous studies, with Weiss et al. reaching an AUC score of up to 0.93 [25], van Schooten et al. reaching an AUC score of 0.82 [19] and Aicha et al. reaching an AUC of 0.72 [1]. Furthermore, Neural Networks consistently performed worse than gradient boosting methods, aligning with prior research [1]. In conclusion, while all three gradient boosting algorithms performed well, CatBoost stood out as a potential

choice for fall risk assessment due to its strong discriminative ability. Our findings outperformed previous research and demonstrated the importance of feature engineering. Integrating these models into embedded systems seems plausible, but practical testing is necessary.

Future research will explore the ability to transfer our current findings to other sensor positions on the body. The current sensor configuration is challenging to wear and obstructs everyday use, presenting a substantial limitation to this research. Adapting the preprocessing and feature engineering strategies for accelerometers placed on alternate body locations, such as the upper back or wrist, could yield valuable insights. The next step involves implementing the findings in a real-world embedded system. In this scenario, the initial step in the preprocessing pipeline involves applying the running window SMA threshold to discard time windows with no activity. Only time windows that surpass a specific activity level will proceed for further calculations. This process is relatively straightforward and can be potentially integrated into a smart sensor, resulting in reduced power consumption. However, reducing the time resolution to ease computational loads is a topic for future research. Ultimately, this research could lead to clinical studies involving individuals wearing the embedded system for fall risk assessment based on our research.

6 Conclusion

Based on the published LTMM dataset, we have developed a pipeline that efficiently categorizes senior individuals as high-risk for falls. First, a preprocessing pipeline was designed and tested to effectively clean sensor data using frequency filter techniques and to recognize walking segments. This preprocessing substantially reduced computational overhead for feature extraction and ML model training, simultaneously enhancing model performance. From the preprocessed data, we computed over 20 distinct features, trained various ML algorithms and examined the impact of features on model outcomes. In our analysis, it became evident that features linked to gait stability held dominant importance, corroborating findings from prior studies [10]. The exploration of different ML models identified gradient boosting algorithms, namely XGBoost, LightGBM, and CatBoost, as top-performing models. Notably, these models exhibited uniform performance across various evaluation metrics, with CatBoost slightly outperforming its counterparts. Conversely, Neural Networks and tree-based methods like Extra Trees or Random Forests yielded notably inferior results. Hence, our research indicates that gradient boosting models are best suited for fall risk prediction using accelerometer data. The computational complexity of the resulting models indicates that they could be effectively integrated into embedded hardware, thereby having the potential to be used in mobile devices. In conclusion, we believe that the methods and insights from this work hold potential for the development of an embedded tool capable of accurately predicting the fall risk of elderly individuals.

Funding. This work is funded by the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen (MKW NRW).

References

1. Aicha, A.N., Englebienne, G., van Schooten, K., Pijnappels, M., Kröse, B.: Deep learning to predict falls in older adults based on daily-life trunk accelerometry. Sensors **18**(5), 1654 (2018). https://doi.org/10.3390/s18051654
2. Bobick, A.F., Johnson, A.Y.: Gait recognition using static, activity-specific parameters. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, p. I. IEEE (2001). https://doi.org/10.1109/CVPR.2001.990506
3. Bohannon, R.W.: Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. Age Ageing **26**(1), 15–19 (1997)
4. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
5. Chidean, M.I., et al.: Full band spectra analysis of gait acceleration signals for peripheral arterial disease patients. Front. Physiol. **9**, 1061 (2018)
6. Dubois, A., Bihl, T., Bresciani, J.P.: Identifying fall risk predictors by monitoring daily activities at home using a depth sensor coupled to machine learning algorithms. Sensors **21**(6) (2021). https://doi.org/10.3390/s21061957
7. Florence, C.S., Bergen, G., Atherly, A., Burns, E., Stevens, J., Drake, C.: Medical costs of fatal and nonfatal falls in older adults: medical costs of falls. J. Am. Geriatr. Soc. **66**(4), 693–698 (2018). https://doi.org/10.1111/jgs.15304
8. Goldberger, A.L., et al.: PhysioBank, PhysioToolkit, and PhysioNet. Circulation **101**(23) (2000). https://doi.org/10.1161/01.cir.101.23.e215
9. Howcroft, J., Kofman, J., Lemaire, E.D.: Prospective fall-risk prediction models for older adults based on wearable sensors. IEEE Trans. Neural Syst. Rehabil. Eng. **25**(10), 1812–1820 (2017). https://doi.org/10.1109/tnsre.2017.2687100
10. Ihlen, E.A.F., Weiss, A., Helbostad, J.L., Hausdorff, J.M.: The discriminant value of phase-dependent local dynamic stability of daily life walking in older adult community-dwelling fallers and nonfallers. Biomed. Res. Int. **2015**, 402596 (2015). https://doi.org/10.1155/2015/402596
11. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. **30** (2017). https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
12. Mathie, M., Coster, A., Lovell, N., Celler, B.: Detection of daily physical activities using a triaxial accelerometer. Med. Biol. Eng. Comput. **41**, 296–301 (2003). https://doi.org/10.1007/BF02348434
13. Moreland, B., Kakara, R., Henry, A.: Trends in nonfatal falls and fall-related injuries among adults aged ≥ 65 years — United States, 2012–2018. MMWR Morb. Mortal. Wkly. Rep. **69**(27), 875–881 (2020). https://doi.org/10.15585/mmwr.mm6927a5
14. Murray, M.P., Drought, A.B., Kory, R.C.: Walking patterns of normal men. JBJS **46**(2), 335–360 (1964)
15. Murray, M.P.: Walking patterns of normal woman. Arch. Phys. Med. Rehabil. **51**, 637–650 (1970)
16. Płońska, A., Płoński, P.: MLJAR: state-of-the-art automated machine learning framework for tabular data. version 0.10.3 (2021). https://github.com/mljar/mljar-supervised

17. Raîche, M., Hébert, R., Prince, F., Corriveau, H.: Screening older adults at risk of falling with the Tinetti balance scale. Lancet **356**(9234), 1001–1002 (2000). https://doi.org/10.1016/S0140-6736(00)02695-7
18. Redfield, M.T., Cagle, J.C., Hafner, B.J., Sanders, J.E.: Classifying prosthetic use via accelerometry in persons with transtibial amputations. J. Rehabil. Res. Dev. **50**(9), 1201–1212 (2013). https://doi.org/10.1682/jrrd.2012.12.0233
19. van Schooten, K.S., Pijnappels, M., Rispens, S.M., Elders, P.J.M., Lips, P., van Dieën, J.H.: Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults. J. Gerontol. A Biol. Sci. Med. Sci. **70**(5), 608–615 (2015). https://doi.org/10.1093/gerona/glu225
20. Shapley, L.S., et al.: A value for n-person games. In: Contributions to the Theory of Games, vol. 2 (1953)
21. Simpson, J.M.: Falls in older people: risk factors and strategies for prevention. Ageing Soc. **21**, 673 (2001)
22. Terroso, M., Rosa, N., Torres Marques, A., Simoes, R.: Physical consequences of falls in the elderly: a literature review from 1995 to 2010. Eur. Rev. Aging Phys. Activ. **11**, 51–59 (2014). https://doi.org/10.1007/s11556-013-0134-8
23. Ullrich, M., et al.: Detection of gait from continuous inertial sensor data using harmonic frequencies. IEEE J. Biomed. Health Inform. **24**(7), 1869–1878 (2020). https://doi.org/10.1109/JBHI.2020.2975361
24. Ullrich, M., et al.: Fall risk prediction in Parkinson's disease using real-world inertial sensor gait data. IEEE J. Biomed. Health Inform. **27**(1), 319–328 (2023). https://doi.org/10.1109/JBHI.2022.3215921
25. Weiss, A., et al.: Does the evaluation of gait quality during daily life provide insight into fall risk? A novel approach using 3-day accelerometer recordings. Neurorehabil. Neural Repair **27**(8), 742–752 (2013). https://doi.org/10.1177/1545968313491004
26. Xiao, W., Lu, Y.: Daily human physical activity recognition based on kernel discriminant analysis and extreme learning machine. Math. Probl. Eng. **2015**, 1–8 (2015). https://doi.org/10.1155/2015/790412