engineering proceedings

MDPI

Proceeding Paper

An Extreme Gradient Boosting Approach for Elderly Falls Classification †

Paulo Monteiro de Carvalho Monson *, Vinicius Toledo Dias, Giovanni de Oliveira de Sousa, Gabriel Augusto David, Fabio Romano Lofrano Dotto and Pedro de Oliveira Conceição Junior

Department of Electrical and Computer Engineering, São Carlos School of Engineering at the University of São Paulo (EESC-USP), São Carlos, SP, Brazil; viniciusdias@usp.br (V.T.D.); giovanni_oliveiradesousa@usp.br (G.d.O.d.S.); gadavid@usp.br (G.A.D.); fabio.dotto@usp.br (F.R.L.D.); pedro.oliveiracjr@usp.br (P.d.O.C.J.)

* Correspondence: paulo.monson@usp.br

† Presented at The 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: Falls pose a significant threat to the elderly population, often resulting in severe health complications such as fractures and other adverse outcomes, which can drastically lower their quality of life. Early detection of fall risks is crucial in mitigating the impact of such events. Various technologies have been developed to address this issue, including alert systems that notify users of imminent risks due to environmental factors or physiological changes. However, accurately detecting and distinguishing between normal activities, imminent fall risk, and actual falls remains a challenge. This study proposes a machine learning approach using the XGBoost algorithm to improve fall detection accuracy among the elderly. A dataset comprising 2039 samples of data on proximity to objects, spatial location changes, heart rate, blood oxygen saturation (SpO2), blood sugar levels, and pressure applied by the user, categorized into normal, imminent fall risk, and fall classes, was utilized to train and test the model. The model was trained on 70% of the data, with 30% allocated for testing. Hyperparameter optimization was performed using a randomized search with cross-validation. Previous studies have reported an accuracy of 0.9667 for the same dataset. In contrast, this study achieved an accuracy of 1.0, demonstrating a significant improvement in overall performance compared to earlier work. The confusion matrix demonstrates the model's ability to distinguish between all three classes with no false positives. Additionally, sensitivity tests were conducted by varying training sample sizes and randomizing data splits, confirming the model's robustness in different conditions. These results show that the proposed method was able to sort correctly all the samples on training and test, outperforming previous studies on detecting fall-related events, reducing the likelihood of false alarms, and enhancing resource allocation for elderly care.

Keywords: smart healthcare; Internet of Medical Things (IoMT); fall detection; elderly falls; fall risk

Citation: de Carvalho Monson, P.M.; Dias, V.T.; de Oliveira de Sousa, G.; David, G.A.; Dotto, F.R.L.; de Oliveira Conceição Junior, P. An Extreme Gradient Boosting Approach for Elderly Falls Classification. *Eng. Proc.* **2024**, *5*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 26 November 2024

1. Introduction

Falls are among the most detrimental events that elderly individuals can experience [15]. These incidents pose a constant threat, especially to older adults whose motor abilities may be significantly impaired after such accidents [12]. According to the World Health Organization (WHO), 30% of individuals over the age of 65 suffer from one or more accidental falls each year, with the incidence increasing to 50% among those aged 80 and older [11]. Furthermore, statistical data indicates that the incidence of falls increased by 31% between 2007 and 2016, with expectations for further growth in the future [3,12]. Most falls do not occur from heights but rather from stumbling or slipping during routine activities like walking [12]. These findings underscore that fall prevention is closely linked to the quality of life for the elderly [17].

The increasing prevalence of falls among independently living older adults highlights the urgent need for proactive solutions [2]. Combined with the accelerating aging of the global population, there is an urgent need to develop more efficient fall detection systems [15] to safeguard and protect the elderly. Fall detection relies on methods that recognize such events based on pattern recognition, which identifies abrupt changes in bodily sensor parameters and detects specific deviations [1,11]. Since older adults are particularly vulnerable to falls due to weakened muscle structures and external conditions, the issue is not only a concern for healthcare professionals but also draws attention from the scientific community and industry. Both sectors are actively working on the development of Fall Detection Systems (FDS) and devices designed to prevent such incidents [5,11].

Therefore, it is imperative to identify solutions that are accurate, reliable, robust, and convenient for the elderly, aiming to mitigate fall occurrences [12]. Studies that monitor and predict falls are crucial in mitigating fall risks, especially when combined with artificial intelligence (AI) techniques and the Internet of Things (IoT). As highlighted by [2], using IoT and AI to detect falls can reduce fall risk by up to 25%, potentially preventing injuries and hospitalizations. Given that the average cost of a fall-related injury for an older adult is approximately $30,000, the use of IoT and AI in fall detection could save healthcare systems billions of dollars annually.

In this context, several solutions have been proposed in the literature. For example, [12] proposed the Calm Stick (cStick), a system that monitors the user's physiological data along with location and environmental information using IoT technology. The system can detect and predict falls with an accuracy close to 95%, based on physiological changes that occur just before a person is about to fall. The device warns the user of an imminent incident and provides a control solution to reduce the impact on the elderly. More recently, [10] introduced a method that combines acceleration and angular velocity data from an Inertial Measurement Unit (IMU) attached to individuals, both with and without fall risk, along with the Gradient Boosting Decision Trees (GBDT) algorithm, to detect falls during the gait cycle.

Similarly, ref. [8] examined the viability of using wearable sensors during walking to identify older adults with early-stage balance issues. Their proposed approach used sensor data placed on the knee and hip of individuals during walking, combined with the Gradient Boosting Machine (GBM) algorithm, to classify individuals into low and high-fall-risk groups. Another method, ref. [2] proposed a fall detection system using smart carpets, employing IoT techniques and deep learning algorithms, including Random Forest (RF), XGBoost, Gated Recurrent Units (GRUs), Logistic Regression (LGR), and K-Nearest Neighbors (KNN).

The methodology proposed by [7] focuses on detecting and classifying falls based on variations in human silhouette shape using computer vision techniques. They utilized multivariate monitoring with an exponentially weighted moving average (MEWMA) and a classification stage based on Support Vector Machines (SVM). Moreover, ref. [14] explored fall detection through image classification and machine learning, analyzing six human activities (walking, sitting, standing, picking up objects, drinking water, and falling) using various algorithms such as random forest, K-nearest neighbors, support vector machines, long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), and convolutional neural networks (CNN), with CNN achieving the highest precision at 95.30%.

Despite these significant advances, as noted by [12], one of the primary challenges in fall classification is the presence of false positives and false negatives, which can trigger unnecessary alerts. Although numerous studies focus on fall detection using individual sensors, such as wearables or depth cameras, the performance of these systems remains suboptimal, particularly due to high false alarm rates [15]. The difficulty in accurate classification stems primarily from individual variability and step-by-step fluctuations, making precise classification challenging [10].

In this context, the main objective of this study is to enhance the predictive capability of systems designed to detect falls among the elderly using data from [12]. Reducing false fall event classification allows for better resource allocation when assisting patients, prevents false alarms, and increases system reliability. This paper presents a model developed using the gradient boosting algorithm known as XGBoost to classify falls in elderly patients.

2. XGBoost

XGBoost, short for eXtreme Gradient Boosting, is an advanced algorithm in ensemble learning that utilizes decision trees. Proposed by Chen and Guestrin [6], it represents an advancement from previous algorithms based on ensembles of decision trees (DT) such as AdaBoost and Gradient Boosting (GB), which integrates several weak prediction models, i.e., DT, to iteratively enhance the model's accuracy in classification or regression tasks. Like GB, XGBoost has an objective function that minimizes the loss function between the predicted and expected response. Notably, it also has a term that regulates the complexity of DT internally [16]. Each DT will produce an output $f_k(x_i)$ and the residuals will be iteratively refined by an objective function. The final output aggregates contributions from all K trees in the ensemble, ensuring robust predictive performance, several DT will be adjusted simultaneously [9].

The prediction response $\hat{p}_i$ of a sample i will be the sum of the responses of each tree according to equation 1, where K is the number of DT and f_k is the function of tree k using the vector of inputs $\boldsymbol{x_i}$.

$$\hat{p}_i = \phi(\boldsymbol{x}) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i) \tag{1}$$

XGBoost minimizes an objective function 2 to optimize its structure. The first component is a sum of the loss function $L(p_i, \hat{p}_i)$ between the predicted value $\hat{p}_i$ and the expected value p_i in n samples, while the second component resolves the sum of the regulatory term $\Omega(f_k)$ according to equation 2 for K trees, where T and ω_j, respectively, are the number of leaves and the score of each leaf j, while γ means the complexity of each leaf and λ is the weight of the regulatory term. This regulatory term is one of the main differences from GB, rewarding a less complex structure and mitigating the risk of overfitting [9].

$$O(\phi) = \sum_{i=1}^{n} L(p_i, \hat{p}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

Key operational choices in XGBoost choices are important to achieve good performance. The γ parameter defines the minimum loss reduction gain for splitting a tree node, i.e., a higher γ value produces less complex trees; other parameters such as tree depth can also reduce the complexity of the model. The loss function can also be chosen by the user, who needs to define the first and second-order gradient. XGBoost also addresses sparse data scenarios in its composition to choose the best division of DT by preliminarily eliminating null or missing values [4].

3. Experimental Analysis

The dataset used in this study is sourced from a previous investigation reported in [12]. It contains 2039 samples comprising data on proximity to objects, spatial location changes, heart rate, blood oxygen saturation (SpO2), blood sugar levels, and pressure applied by the user. The samples are divided into three distinct classes: normal (n = 690), imminent fall risk (n = 682) and falling (n = 667).

The computational model for detecting falls was developed using the XGBoost algorithm. Pre-processing, training, and testing of the model were all conducted in Python.

The dataset was split into a training set comprising 70% of the samples (n = 1427) and a test set containing the remaining 30% (n = 612). This split was stratified to preserve the proportion of classes across both sets. To ensure reproducibility, a random seed of 25 was used during the dataset split and model optimization. The XGBoost algorithm, by its nature, does not require feature normalization, so no additional pre-processing steps were applied.

4. Results and Discussion

An initial exploratory data analysis was conducted to examine the distribution and variance of key features in both the training and test sets. Table 1 summarizes the mean and standard deviation of the main features for each of the three classes (normal, imminent fall risk, and fall).

Hyperparameter optimization was performed using the RandomizedSearchCV method from the Scikit-learn library. This optimization used 1000 iterations with 10-fold cross-validation. Table 2 presents the hyperparameter search space and the optimal values obtained after tuning.

Table 1. Mean and standard deviation of training and test sets.

Feature	Mean	Std	Class
Distance	60.05 (59.843)	5.864 (5.589)	
HRV	75.25 (74.933)	8.893 (8.48)	Normal
Sugar Level	75.194 (75.086)	3.024 (2.882)	n = 483 (207)
SpO2	95.194 (95.086)	3.023 (2.881)	
Distance	20.189 (20.259)	6.005 (5.668)	
HRV	97.475 (97.527)	4.409 (4.166)	Imminent fall risk
Sugar Level	50.377 (50.517)	12.010 (11.337)	n = 477 (205)
SpO2	85.094 (85.129)	3.002 (2.834)	
Distance	4.992 (5.002)	2.884 (2.911)	
HRV	114.983 (115.005)	5.769 (5.822)	Fall
Sugar Level	93.103 (94.676)	75.039 (75.417)	n = 467 (200)
SpO2	69.983 (70.004)	5.769 (5.823)	

Table 2. Hyperparameter space for optimization of the XGBoost algorithm.

Parameter	Distribution	Lower Bound	Upper Bound	Optimum Value
colsample_bytree	Uniform	0.7	1	0.961
gamma	Uniform	0.0	0.5	0.291
learning_rate	Uniform	0.0003	0.3	0.084
subsample	Uniform	0.6	1	0.764
max_depth	Random Int.	2	6	5
n_estimators	Random Int.	100	150	123

The optimal hyperparameters obtained (‘colsample_bytree’: 0.961, ‘gamma’: 0.291, ‘learning_rate’: 0.084, ‘max_depth’: 5, ‘n_estimators’: 123, ‘subsample’: 0.764) were adopted for training the model. The model’s performance was evaluated using the test set, and its results are illustrated in the confusion matrix shown in Figure 1. The matrix reveals the model’s high classification accuracy across the three classes: normal, imminent fall risk, and fall, with no false positives or negatives. This outcome translates to an overall accuracy of 100%, exceeding the accuracy of 96.67% reported in [12].

To further evaluate the model’s robustness and minimize the risk of overfitting, additional tests were conducted by varying the number of training samples (from 60% to

90%) and removing the random seed to allow for different training/test splits. Across these various configurations, the model consistently maintained its high performance, confirming its robustness and generalizability.

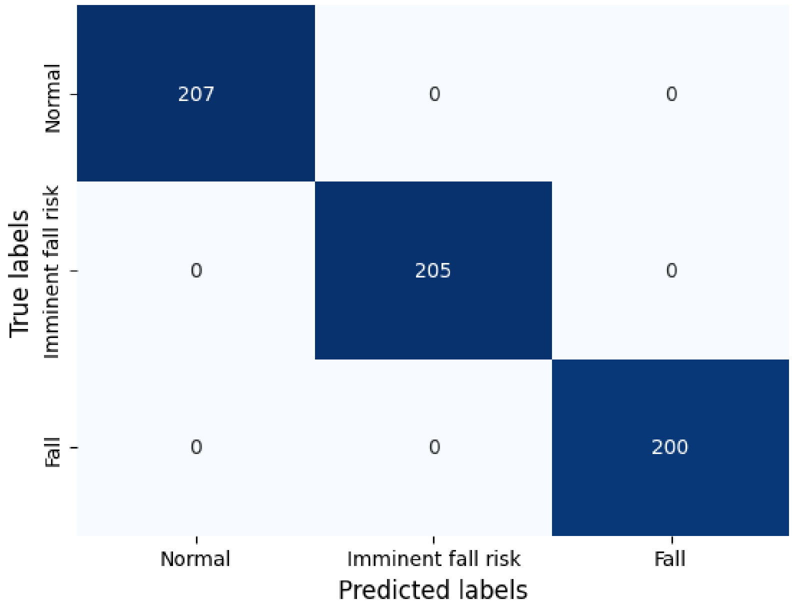

Figure 1. Confusion matrix for test data.

Additionally, to corroborate the precision of the model, we included a principal component analysis (PCA) visualization shown in Figure 2. This analysis demonstrates that the separation of the three classes—normal, imminent fall risk, and fall—is well-defined, further supporting the high accuracy observed in the confusion matrix. The explained variance ratio for the first two principal components is 0.97778, indicating that these two components capture most of the variance in the data, reinforcing the distinct separation of the classes.

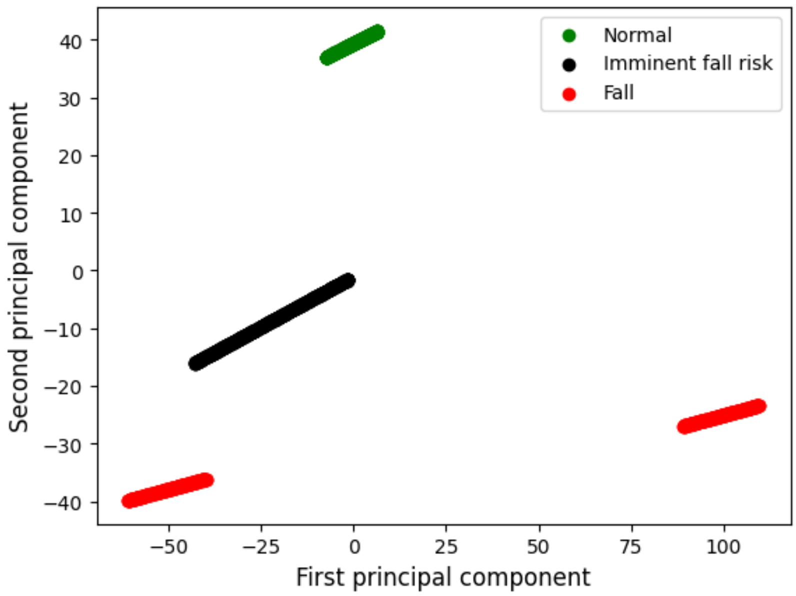

Figure 2. Principal component analysis for test data.

5. Conclusions

This study presents a robust method for detecting potential fall events among the elderly, contributing to a significant improvement in prediction accuracy over previous approaches. The proposed method leverages the XGBoost algorithm, optimized through hyperparameter tuning, and achieves high precision, distinguishing all data between normal activity, imminent risks of fall, and actual falls. The reduction in false alerts is particularly noteworthy, as it supports a more efficient allocation of care resources for elderly patients, ensuring that help is provided when truly necessary. The method's effectiveness was confirmed through testing with different configurations, establishing its potential for actual patient applications. Future work may involve integrating this method into real-time monitoring systems for further validation and exploring its adaptability to other datasets.

Author Contributions: Conceptualization, P.M.d.C.M. and G.A.D.; methodology, V.T.D., F.R.L.D. and P.d.O.C.J.; software, P.M.d.C.M. and G.d.O.d.S.; validation, G.A.D. and V.T.D.; formal analysis, P.M.d.C.M., V.T.D., G.A.D. and P.M.d.C.M.; investigation, P.M.d.C.M., V.T.D. and G.d.O.d.S.; resources, P.d.O.C.J.; data curation, P.M.d.C.M. and G.d.O.d.S.; writing—original draft preparation, P.M.d.C.M., V.T.D. and G.A.D.; writing—review and editing, P.d.O.C.J. and F.R.L.D.; visualization, G.d.O.d.S.; supervision, P.d.O.C.J.; project administration, P.d.O.C.J.; funding acquisition, P.d.O.C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES)—Finance Code 001, in part by the National Council for Scientific and Technological Development (CNPq), grant #140775/2024-2, and by the Pro-Rectory of Research and Innovation of the University of São Paulo under grant: #22.1.09345.01.2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the National Council for Scientific and Technological Development (CNPq), and the Pro-Rectory of Research and Innovation of the University of São Paulo for the financial support of this research.

Data Availability Statement: Publicly available dataset was analyzed in this study. This data can be found here: https://www.kaggle.com/datasets/laavanya/elderly-fall-prediction-and-detection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alanazi, T.; Babutain, K.; Muhammad, G. A Robust and Automated Vision-Based Human Fall Detection System Using 3D Multi-Stream CNNs with an Image Fusion Technique. *Appl. Sci.* **2023**, *13*, 6916. https://doi.org/10.3390/app13126916.
2. Alharbi, H.A.; Alharbi, K.K.; Hassan, C.A.U. Enhancing Elderly Fall Detection through IoT-Enabled Smart Flooring and AI for Independent Living Sustainability. *Sustainability* **2023**, *15*, 15695. https://doi.org/10.3390/su152215695.
3. Avin, K.G.; Hanke, T.A.; Kirk-Sanchez, N.; Mcdonough, C.M.; Shubert, T.E.; Hardage, J.; Hartley, G. Management of Falls in Community-Dwelling Older Adults: Clinical Guidance Statement from the Academy of Geriatric Physical Therapy of the American Physical Therapy Association. *Phys. Ther.* **2015**, *95*, 815–834. https://doi.org/10.2522/ptj.20140415.
4. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.
5. Chaccour, K.; Darazi, R.; El Hassani, A.H.; ANDRÈS, E. From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems. *IEEE Sens. J.* **2017**, *17*, 812–822. https://doi.org/10.1109/JSEN.2016.2628099.
6. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. https://doi.org/10.1145/2939672.2939785.
7. Harrou, F.; Zerrouki, N.; Sun, Y.; Houacine, A. Vision-based fall detection system for improving safety of elderly people. *IEEE Instrum. Meas. Mag.* **2017**, *20*, 49–55. https://doi.org/10.1109/MIM.2017.8121952.
8. Hu, Y.; Bishnoi, A.; Kaur, R.; Sowers, R.; Hernandez, M.E. Exploration of Machine Learning to Identify Community Dwelling Older Adults with Balance Dysfunction Using Short Duration Accelerometer Data. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 812–815. https://doi.org/10.1109/EMBC44109.2020.9175871.

9. Niazkar, M.; Menapace, A.; Brentan, B.; Piraei, R.; Jimenez, D.; Dhawan, P.; Righetti, M. Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). *Environ. Model. Softw.* **2024**, *174*, 105971. https://doi.org/https://doi.org/10.1016/j.envsoft.2024.105971.
10. Nishiyama, D.; Arita, S.; Fukui, D.; Yamanaka, M.; Yamada, H. Accurate fall risk classification in elderly using one gait cycle data and machine learning. *Clin. Biomech.* **2024**, *115*, 106262. https://doi.org/https://doi.org/10.1016/j.clinbiomech.2024.106262.
11. Purwar, A.; Chawla, I. A systematic review on fall detection systems for elderly healthcare. *Multimed. Tools Appl.* **2024**, *83*, 43277–43302. https://doi.org/10.1007/s11042-023-17190-z.
12. Rachakonda, L.; Mohanty, S.P.; Kougianos, E. cStick: A Calm Stick for Fall Prediction, Detection and Control in the IoMT Framework. *In Proceedings of the Internet of Things. Technology and Applications, 4th IFIP International Cross-Domain Conference, IFIPIoT 2021, Virtual Event, 4–5 November 2021*; Camarinha-Matos, L.M., Heijenk, G., Katkoori, S., Strous, L., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 129–145.
13. Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. *IEEE Trans. Circuits Syst. Video Technol.* **2011**, *21*, 611–622. https://doi.org/10.1109/TCSVT.2011.2129370.
14. Taylor, W.; Dashtipour, K.; Shah, S.A.; Hussain, A.; Abbasi, Q.H.; Imran, M.A. Radar Sensing for Activity Classification in Elderly People Exploiting Micro-Doppler Signatures Using Machine Learning. *Sensors* **2021**, *21*, 3881. https://doi.org/10.3390/s21113881.
15. Wang, X.; Ellul, J.; Azzopardi, G. Elderly fall detection systems: A literature survey. *Front. Robot. AI* **2020,** *7, 71*.
16. Zhang, J.; Wang, R.; Lu, Y.; Huang, J. Prediction of Compressive Strength of Geopolymer Concrete Landscape Design: Application of the Novel Hybrid RF–GWO–XGBoost Algorithm. *Buildings* **2024**, *14*, 591. https://doi.org/10.3390/buildings14030591.
17. Zhang, S.; Xu, W.; Zhu, Y.; Tian, E.; Kong, W. Impaired multisensory integration predisposes the elderly people to fall: A systematic review. *Front. Neurosci.* **2020**, *14*, 411.