Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/99/20251790

Research on the Elderly Fitness Fall Risk Prediction Model Based on Decision Tree

Jinxiu Han[1,a], Jiayue Huang[1,b], Yutong Luo[2,c], Ping Yang[1,d,*]

[1]School of Foreign Languages, Zhuhai College of Science and Technology, China
[2]School of Business, Zhuhai College of Science and Technology, China
a. 18050702759@163.com, b. 2671104571@qq.com, c. 3181530800@qq.com,
d. jasmin125@zcst.edu.cn
**corresponding author*

Abstract: As the global aging process intensifies, the issue of falls has become a common health risk in the daily lives of the elderly, especially in physical fitness activities. This article establishes a fall risk prediction model using a decision tree algorithm based on fitness data of the elderly and analyzes the relationship between various health factors and fall risk. The study found that features such as heart rate variability, blood sugar levels, and exercise duration have significant impacts on fall prediction. Through the analysis of the ROC curve and PR curve of the model, the results show that the model performs well in the prediction accuracy and risk identification of fall events. The research in this article not only provides a predictive tool for the safety of the elderly in fitness but also provides a theoretical basis and practical guidance for preventing falls in the elderly and formulating personalized safe exercise plans.

Keywords: Decision Tree, the aged, Gym fitness, fall down.

1. Introduction

The risk of falls is particularly high during physical exercise and daily activities, as routine movements for younger individuals become hazardous for older adults due to declines in balance, muscle strength, and reaction time. Simple tasks like walking or climbing stairs pose significant challenges, and the physical exertion during exercise can increase the likelihood of falls, as elderly individuals may struggle to maintain balance and stability.

Additionally, falls have broader implications for families and society, often necessitating medical treatment, hospitalization, or long-term care, which burdens healthcare systems and caregivers. Prolonged recovery can strain families emotionally and financially. The societal costs include increased demand for medical resources and rehabilitation services, highlighting the urgent need for preventive measures like fall detection systems, balance training, and safer living environments.

This article proposes a fall prediction model using a decision tree algorithm based on fitness data for the elderly, analyzing health indicators such as muscle strength, exercise duration, heart rate variability, and blood sugar levels in relation to fall events. The goal is to predict fall risk during physical activities, offering theoretical foundations and practical guidance for fitness safety management, ultimately aiming to reduce falls, improve quality of life, and alleviate medical burdens on families and society.

2. Literature review

Fall detection and prevention is a crucial area in healthcare, especially concerning the elderly population. Several studies have utilized machine learning to detect and prevent falls effectively.

The issue of falls among the elderly and its severe consequences has been extensively studied. Kannus et al. noted that falls are one of the leading causes of injuries and deaths among the elderly [1]. Rubenstein further explored the epidemiology, risk factors, and prevention strategies of falls, emphasizing the importance of preventive measures in reducing fall incidents [2]. These studies indicate that falls not only significantly impact the quality of life of the elderly but also pose a burden on the healthcare system.

A systematic review by Rastogi and Singh [3] looked into fall detection systems using machine learning. The authors explored different machine learning techniques used in detecting falls and assessed their performance. In a conference paper by Vallabh et al. [4], they used machine learning algorithms to detect falls. Their study demonstrated the potential of machine learning in fall detection, showing that these algorithms could accurately detect falls in real-time.

Usmani et al. [5] conducted a systematic review of the latest research trends in fall detection and prevention using machine learning. They found that recent advancements in machine learning technologies have significantly improved fall detection systems' performance. They also noted that these advancements have led to new preventive measures and interventions, minimizing the risk of falls. Yacchirema et al. [6] focused on using IoT and ensemble machine learning algorithms for fall detection in the elderly. The authors developed a system that integrated sensors and machine learning algorithms to detect falls accurately. Their study showed that combining IoT and machine learning could provide an effective and efficient solution to fall detection. Thakur and Han [7] studied fall detection in assisted living and aimed to identify and improve the optimal machine learning method. The authors found that specific machine learning methods could significantly improve the accuracy and effectiveness of fall detection in assisted living environments.

Lastly, Chelli and Pätzold [8] proposed a machine learning approach for fall detection and daily living activity recognition. They developed a system that could distinguish between normal daily activities and falls, providing a more accurate and reliable fall detection system.

Machine learning plays an important role in fall detection and prevention. While different studies have proposed various machine learning techniques, overall, they all contribute to improving fall detection systems' accuracy and efficiency. Future research should continue to explore and optimize machine learning methods for fall detection and prevention as it can significantly impact healthcare, particularly for the elderly population.

3. Data and Model

3.1. Data and brief analysis

The data set used in this study includes seven variables: Muscle Strength, Exercise Duration HRV, Heart Rate Variability (HRV), Sugar Level, oxygen saturation (SpO2), Posture and Decision. These variables were selected based on their correlation with the fall risk of the elderly during exercise. Muscle strength measures the muscle ability of participants, which is the basic factor to maintain stability and carry out daily activities; The duration of exercise reflects the duration of exercise, and too long exercise time may lead to fatigue, which in turn affects the balance and coordination ability; Heart rate variability represents the function of autonomic nervous system. Low heart rate variability may mean that participants' ability to adjust their balance in emergency is weakened, thus increasing the risk of falling. Blood sugar level measures the blood sugar state before exercise, and too low or too high blood sugar will have an impact on physical function; Blood oxygen saturation reflects the

DOI: 10.54254/2755-2721/99/20251790

blood oxygen supply of the participants, and low blood oxygen saturation may lead to listlessness and muscle weakness; Body posture describes the performance of participants in the process of exercise, and incorrect posture may affect the stability and center of gravity control of the body. These characteristic variables are commonly used to analyze the possibility of falls of the elderly during exercise, and the Decision variable is used as the target variable to indicate whether a fall event occurs. By analyzing the relationship between these characteristic variables and target variables, the research aims to establish an effective prediction model and provide reference for the elderly to exercise safely.

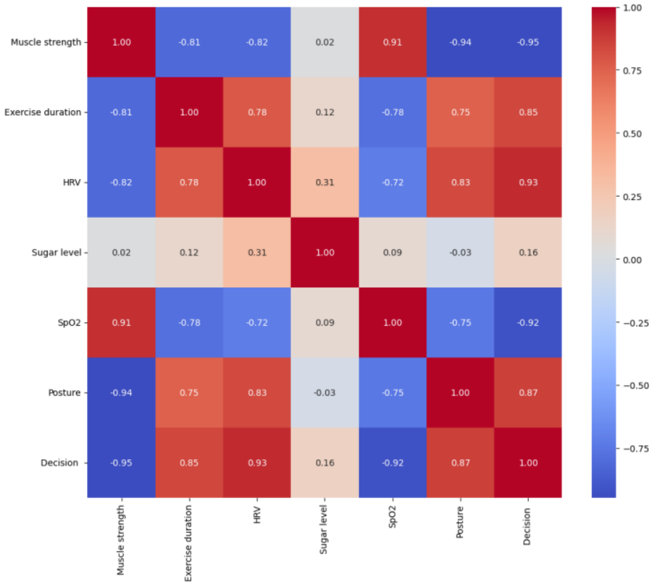

Figure 1: Correlation tables of selected variables

According to the correlation analysis heat map, it can be found that there are significant correlations among multiple variables. For example, the correlation coefficient between "Muscle Strength" and "Decision" is -0.946120, indicating that there is a strong negative correlation between them, that is, the stronger the muscle strength, the lower the possibility of falling. Similarly, the correlation coefficient between "muscle strength" and "Posture" is -0.941711, indicating that the stronger the muscle strength, the more stable the posture. In addition, the correlation coefficient between "Exercise Duration" and "decision-making" is 0.847559, indicating that the longer the exercise time, the higher the possibility of falling. The correlation between "Sugar Level" and other variables is relatively weak, which may mean that the blood sugar level has little effect on the possibility of falling. To sum up, muscle strength, exercise duration and posture may be important factors that affect the risk of falls during exercise for the elderly.

3.2. Decision tree model

A decision tree classification model is a machine learning algorithm that classifies data into categories using a tree-like structure of decision rules. It starts with a root node representing the entire dataset and splits data at each node based on features that maximize class distinction, using metrics like Gini impurity or information gain. The process continues until leaf nodes are reached, where final classifications are made. Decision trees are interpretable and easy to visualize, but they can overfit the data if overly complex. Techniques like pruning and ensemble methods, such as random forests, can help improve generalization and reduce overfitting.

DOI: 10.54254/2755-2721/99/20251790

3.3. Model evaluation index

The quality of a decision tree classification model is typically evaluated using several key criteria, ensuring that the model performs well and generalizes effectively to new data.

Accuracy: This is one of the most straightforward metrics, measuring the percentage of correct classifications made by the model on a test dataset. High accuracy indicates that the decision tree is effectively separating classes based on the features it was trained on.

Precision and Recall: These metrics become particularly important in cases of class imbalance. Precision measures how many of the positive predictions were actually correct, while recall (or sensitivity) measures how well the model identifies all positive cases. A good decision tree should balance these two, ensuring it captures the majority of positive instances without introducing too many false positives.

F1 Score: This combines precision and recall into a single metric, offering a balanced measure that is useful when there's a trade-off between these two. A high F1 score indicates that the model performs well in both identifying the true positives and minimizing false positives.

AUC-ROC Curve: The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve provides insight into the model's ability to differentiate between classes. A higher AUC means the model has strong discriminatory power, with an ideal value close to 1, indicating it can effectively separate positive from negative classes.

Overfitting and Generalization: A good decision tree model should generalize well to new, unseen data. Overfitting occurs when the tree becomes too complex, capturing noise and nuances from the training data that don't apply to other datasets. To avoid overfitting, techniques like pruning can be used, where branches that have little predictive power are removed.

4. Decision tree model results

Figure 2 illustrates that the model demonstrates strong classification capabilities, particularly in distinguishing between the positive and negative classes with a high degree of accuracy. The area under the curve (AUC) is measured at 0.81, signifying that the model has an 81% probability of correctly ranking positive samples before negative ones. This high AUC score highlights the model's effective ability in predicting falls, as it suggests a strong discriminatory power between the two classes.

Typically, an AUC value between 0.8 and 0.9 indicates that the model is performing well, making reliable predictions with a good balance between sensitivity and specificity. Moreover, the shape of the receiver operating characteristic (ROC) curve further supports this conclusion. The curve closely approaches the upper left corner of the graph, which is the ideal position for a classifier. This proximity to the upper left corner suggests that the model achieves a low false positive rate while maintaining a high true positive rate. In essence, the model is not only accurate in identifying the positive class but also efficient in minimizing the incorrect classification of negative samples, reinforcing its robustness in fall prediction tasks.

DOI: 10.54254/2755-2721/99/20251790

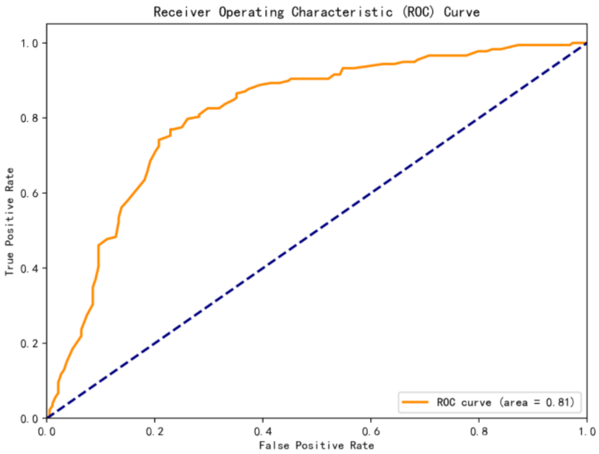

Figure 2: ROC Curve

Figure 3 indicates that while the model exhibits a reasonable degree of accuracy in predicting positive classes, there is still room for improvement in controlling false positives. The area under the curve (AUC) for this model is 0.75, which suggests that although the model performs well in identifying positive cases, it occasionally misclassifies negative samples as positives, leading to some false positives.

AUC values around 0.75 suggest moderate classification performance, but the presence of false positives highlights the need for optimization, especially in scenarios where reducing such errors is critical. In cases of data imbalance, precision-recall (PR) curves tend to offer more insightful evaluations of model performance compared to ROC curves. PR curves emphasize the trade-off between precision (the proportion of true positives among all predicted positives) and recall (the proportion of true positives identified out of all actual positives), making them particularly useful when the dataset has an unequal distribution of classes.

In this model's PR curve, the relatively flat shape implies that there is a balance point where precision and recall are well-aligned. However, the noticeable drop in precision at the tail end of the curve indicates that as the recall rate increases, the precision rate diminishes, suggesting a growing number of false positives. This trade-off implies that the model may be overly eager to capture as many positive cases as possible, but at the cost of sacrificing precision, thereby introducing more errors in the form of false positives. Consequently, while the model is reasonably effective, improvements could be made in reducing the false positive rate to enhance its reliability, particularly in applications where precision is critical.

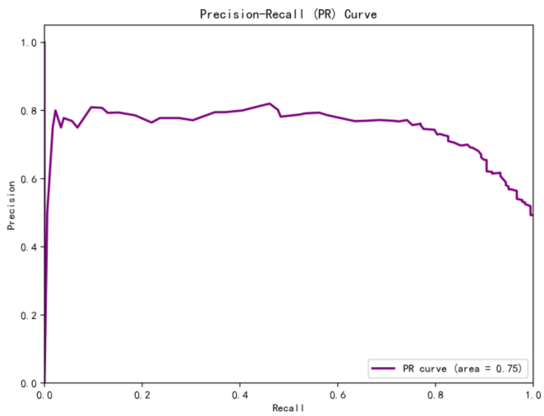

Figure 3: PR Curve

The results of fig. 4 are analyzed as follows. The root node in the figure is heart rate variability (HRV ≤ 105.021). This variable is used to preliminarily distinguish whether there is a risk of falling. Lower heart rate variability (HRV) may mean that the participants' balance ability is decreased, so there is a higher risk of falling.

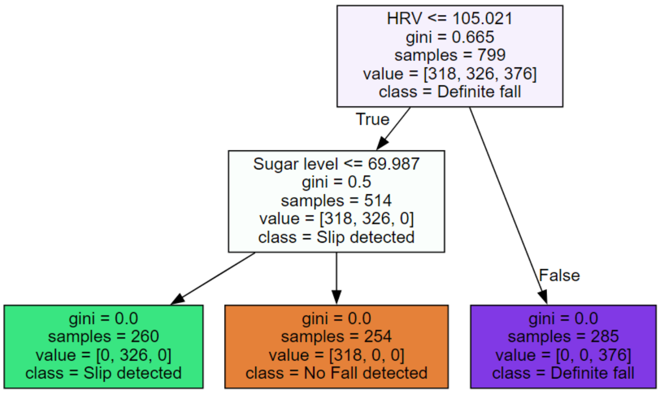

Figure 4: Decision tree classification structure

Left subtree (True branch): If HRV ≤ 105.021, the classification is further based on "Sugar Level ≤ 69.987". The influence of blood sugar level on falls may not be as strong as other variables in your data, but it is still an important indicator.

Left child node: If the Sugar Level ≤ 69.987, it will be displayed as "Slip Detected". The Gini coefficient here is 0, which means that the classification is completely correct, and all 260 samples belong to the "slip detection" category.

Right child node: if the sugar level is greater than 69.987, it is judged as "No Fall Detected". This branch also shows that the classification is completely correct, and all 254 samples belong to the "no fall" category.

Right subtree (False branch): if HRV > 105.021, it is directly judged as "Definite Fall", and all 285 samples belong to this category.

For the indexes involved, Gini coefficient is an index to measure the impurity of classification, and the lower the value, the higher the purity. At the final leaf node, Gini coefficient is 0, which indicates that the categories of samples are completely consistent, that is, the classification of these data by the model is perfect.

5. Conclusion and enlightenment

This study analyzed the relationship between various health indicators—such as heart rate variability, blood sugar level, and exercise time—and fall events in the elderly by constructing a decision tree-based prediction model. The results indicate that monitoring these key health indicators effectively predicts fall risk. Key findings include:

The decision tree model demonstrated good performance, with an AUC of 0.81 (ROC) and 0.75 (PR), indicating strong predictive capabilities, particularly in distinguishing between fall events and non-events. The ROC curve confirms high overall prediction accuracy, while the PR curve shows good accuracy in predicting falls but suggests room for improvement regarding false positives.

Key features influencing fall risk include heart rate variability and blood sugar levels. A heart rate variability below 105.021 significantly increases fall risk, and low blood sugar levels (≤69.987) also heighten the likelihood of falls. These findings indicate that declines in autonomic nervous system function and blood sugar fluctuations are significant risk factors for falls in older adults.

DOI: 10.54254/2755-2721/99/20251790

References

[1] Kannus P, Parkkari J, Koskinen S, et al. Fall-induced injuries and deaths among older adults[J]. Jama, 1999, 281(20): 1895-1899.

[2] Rubenstein, L. Z. (2006). Falls in older people: epidemiology, risk factors and strategies for prevention. Age and Ageing, 35(suppl_2), ii37-ii41. DOI: 10.1093/ageing/afl084

[3] Rastogi S, Singh J. A systematic review on machine learning for fall detection system[J]. Computational intelligence, 2021, 37(2): 951-974.

[4] Vallabh P, Malekian R, Ye N, et al. Fall detection using machine learning algorithms[C]//2016 24th international conference on software, telecommunications and computer networks (SoftCOM). IEEE, 2016: 1-9.

[5] Usmani S, Saboor A, Haris M, et al. Latest research trends in fall detection and prevention using machine learning: A systematic review[J]. Sensors, 2021, 21(15): 5134.

[6] Yacchirema D, de Puga J S, Palau C, et al. Fall detection system for elderly people using IoT and ensemble machine learning algorithm[J]. Personal and Ubiquitous Computing, 2019, 23(5): 801-817.

[7] Thakur N, Han C Y. A study of fall detection in assisted living: Identifying and improving the optimal machine learning method[J]. Journal of sensor and actuator networks, 2021, 10(3): 39.

[8] Chelli A, Pätzold M. A machine learning approach for fall detection and daily living activity recognition[J]. IEEE Access, 2019, 7: 38670-38687.