ELDERLY FALL RISK PREDICTION USING CYCLOSTATIONARY AND TIME-DOMAIN FEATURES

Reem BROME[1], Jad NASREDDINE[2], Frédéric BONNARDOT[1], Mohamed EL BADAOUI[1], Mohamad O. DIAB[3]

[1]Univ Lyon, UJM Saint Etienne, LASPI,
EA 3059, F-42334, IUT de Roanne, FRANCE

[2]i2CAT Foundation, C Gran Capità 2-4
Edifici Nexus I 2ª planta 08034 Barcelona, SPAIN

[3]Rafik Hariri University, Computer and Information Systems Department,
Meshref, LEBANON

reem.abou.marak.dit.brome@univ-st-etienne.fr, jad.nasreddine@i2cat.net
frederic.bonnardot@univ-st-etienne.fr, Mohamed.elbadaoui@univ-st-etienne.fr,
diabmo@rhu.edu.lb

Résumé – Les chutes sont une cause de mortalité importante chez les personnes âgées, il est donc nécessaire d'être capable d'identifier les personnes susceptibles de chuter ("chuteurs") afin de mettre en place d'éventuelles mesures préventives. Nous proposons ici d'évaluer ce risque à l'aide d'une intelligence artificielle entrainée par une base de données constituée de 520 sujets "chuteurs" et "non-chuteurs". Cette base contient pour chaque sujet des mesures de pression réalisées l'aide de semelles instrumentées dans trois conditions de marche en plus de données personnelles. Ces mesures sont traitées afin de construire des indicateurs (tirant notamment parti de la cyclostationarité). Les meilleurs indicateurs sont ensuites sélectionnés afin de tester les performances de plusieurs algorithmes de classification. Finalement, nous validons l'apport de la cyclostationnarité et retenons l'algorithme ANN qui donne ici meilleurs résultats (précision de 89,81 %).

Abstract – Falling is associated with severe morbidity among the elderly community. Therefore, utilizing elderly fall-risk prediction models to predict future cases of falling is essential as a preventive approach. This study aims to optimize and improve the performance of supervised machine-learning models to classify 520 elderly subjects as fallers or non-fallers during 3 different walking conditions. The features used for building the models were extracted from the pressure signals of the innersoles of the subjects. The feature set includes time-domain and cyclostationary features. In addition, two different types of feature selection methods were used and compared. Our study showed that the highest accuracy of 89.81% was achieved using ANN combined with Sequential Backward Selection. The results also help in identifying the important features for detecting elderly people with the risk of falling.

1 Introduction

The common causes and risk factors of falls in older adults were investigated in [1]. The results of this study revealed that some of these risk factors include being 80 years old or above, having muscle defects, previous falls, taking multiple medications, the use of assistive devices, and impairments in stride movements. It was also mentioned that the significance of falls within the elderly community is not just limited to the point that the frequency of the number of falls increases with age, but also that the severity of the injury is highest among the subjects with a history of multiple prior falls, leading to an increase in medical services and rehabilitation expenses. Therefore, reducing the risk of elderly falls is vital from a social and economic point of view. Hence, implementing prevention strategies should emphasize education, training, building safer environments, establishing effective policies to lower susceptibility, and encouraging elderly fall-related research [1]. Thus, there is growing interest in predicting future elderly falls to help reduce their risk of occurrence. In [2], the prediction of elderly fallers was studied using machine-learning algorithms where their best classification model achieved 65% accuracy and 59% sensitivity, with Relief-F feature selection method, and using pressure-sensing-insole and left-shank-accelerometer as predictors [2]. Properties of cyclostationarity in gait signals have been used in modeling and analyzing human walk and ground reaction force (GRF) signals [3] [4]. In [3] an alternative framework for studying GRF signals was proposed based on cyclostationary characteristics rather than the traditional signal processing methods, which assume statistically stationary signal components. In [4], the Cyclostationary (CS) properties and features such as the cyclic autocorrelation function were examined and exploited. Their work demonstrated a significant difference in the cyclic autocorrelation of fallers and non-fallers [4]. One indicator of cyclostationarity is the degree of cyclostationarity (DS) [5]. Using the average degree of cyclostationarity as a single feature in classification models showed promising results in [6] with 68.43% accuracy using the K-nearest neighbors method.

This paper is an extension of our work where we explore some of the best features for the prediction of older adults at risk of falling were the stride time in different walking conditions, the degree of cyclostationariy, and gender in walking conditions involving secondary tasks. Combining the features of all types of walking conditions in an Artificial Neural Network (ANN) classification model and using the Relief-F feature selection method led to improved performance in terms of accuracy that reached 81.16%[7]. In this paper we introduce a new feature, which is the average difference in pressure between toes and heels, use two feature selection methods, and use grid search cross-validation as a hyperparameter tuning method to improve and optimize the classification model.

The paper is organized as follows. Section 2 explains our collected dataset, experimental design, and describes the classification models. More specifically, it presents the data collection process, feature extraction and selection, and the used classification methods.In Section 3, we present and discuss the performance of the studied models. In section 4 we finally summarize our findings, limitations, and prospects.

2 Model Description

2.1 Database Description

The database used in this study is from the original series of the study by the LPE (Laboratoire de Physiologie de l'Exercice) and CHU (Centre Hospitalo-Universitaire) of Jean Monnet St-Etienne University. The description of the setup can be found in [6] and in figure 1. The system was designed to record four independent pressure signals : left heel, left toes, right heel, and right toes. Elderly patients were recruited to participate in this experiment and they were instructed to walk wearing these sensors for 20 meters in a straight line. After the test trial, each participant was asked to walk this distance three times. The first time is the baseline where they walked without performing secondary tasks (MS). The second time, they walked while enumerating aloud as many animal names as they could remember (MF). The third time, they walked the same distance again but while de-counting from 50 (MD).

520 healthy elderly patients were recruited for building this database. Their age was 78$\pm$1.08. Out of the 520 subjects, 302 were females, and 217 were males. Only 54 reported that they had previous falls in the past while the rest reported that they had not. As a first stage working with this largely unbalanced data, we included in our study the 54 fallers and randomly chose 54 non-fallers to build classification models that can accurately classify fallers and non-fallers.

2.2 Features Extraction

After conducting further statistical and visual analysis on the dataset in [6] [7], we were able to identify the features listed in

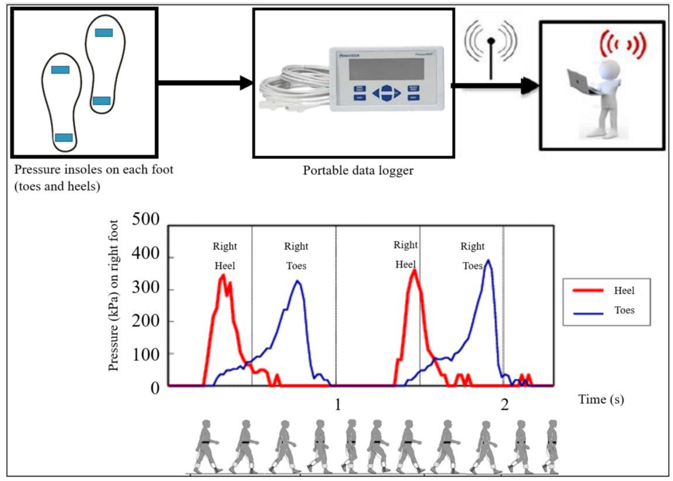

FIGURE 1 – The sensor system and pressure signal recorded.

TABLE 1 – Features for Each Type of Walking Condition : MS, MF, and MD

Abbreviation	Description
PW_R	Pulse Width of the Right Foot
US_R	Undershoot of the Right Foot
US_L	Undershoot of the Left Foot
DTC_L	Duty Cycle of the Left Foot
SR_R	Slew Rate of the Right Foot
SR_L	Slew Rate of the Left Foot
Range_R	Range of the Right Foot
Range_L	Range of the Left Foot
Skw_R	Skewness of the Right Foot
Skw_L	Skewness of the Left Foot
Gender	Male or Female
M_ST	Mean of the Stride Time
STD_ST	Standard Deviation of the Stride Time
Diff_P	Difference in Pressure in Toes and Heels
DC	Degree of Cyclostationarity

table 1 for the three types of walks : MS, MF, and MD. For each type of walk, we have 14 features (a total of 42 features) in addition to the 'Gender' feature. Therefore, there is a total of 43 features.

2.3 Feature Selection Algorithms

The motivation behind using feature selection methods is mainly to simplify models, reduce training duration, improve computational efficiency, and reduce model's generalization error by discarding irrelevant features or noise. This paper utilizes two different feature selection methods : Relief-F and Sequential Backward Selection (SBS).

Relief-F is a filter type selection algorithm that measures feature importance and ranks them based on feature relevance

TABLE 2 – Hyperparameters Tuned Using Grid Search CV

Algorithm	Hyperparameters Chosen
KNN	Leaf size= 5, K=7, Euclidean distance
SVM	Gaussian RBF, Kernel, C(regularization)= 10, Gamma=0.1
ANN	number of neurons=30, activation function= Sigmoid, optimizer, learning rate=0.07, batch size=9, and epochs =50
Decision Trees	criterion= gini, splitter=best, Min samples required to split= 2, min samples required to be at leaf node = 1
Logistic Regression	C=0.3 and Alpha=0.2

to the response, regardless of the classification method used. Relief-F ranks feature based on the identification of feature value differences between nearest-neighbor instance pairs. The feature score decreases if a feature value difference is observed in a neighboring instance pair with the same class. On the other hand, the feature score increases if a feature value difference is observed in a neighboring instance pair with different class values[10].

Then SBS, on the other hand, is a wrapper-type feature selection algorithm where the selection criterion measures the change in classification model performance(e.g., accuracy) that results from removing a feature. The algorithm works by sequentially removing features from all the features to reach the best subset of features. At each stage of removal, the feature that causes the minor performance loss gets excluded [10].

2.4 Classification Algorithms

Five classification methods are used in this study : K-Nearest Neighbors, Support Vectors Machines, Artificial Neural Networks, Decision Trees, and Logistic Regression. In addition, automated hyperparameter tuning using grid search cross validation was performed using Scikit-Optimize Library in python [8] for each classification model to determine the specific model configuration arguments that result in the best classification performance on the given dataset. The obtained hyperparameters for each classification algorithm are listed in table 2.

3 Results and Discussion

A 100 times 10 folds cross-validation was applied on all classifier models. For each classification algorithm 3 different feature sets from all walking conditions were used : all features, features selected by Relief-F, and features selected by SBS.

The features selected by each feature selection technique are shown in table 3. It is noticeable that the common most important features include : the difference in pressure between toes and heel, mean stride time, standard deviation of stride time, and undershoot of left foot, degree of Cyclostationarity and gender. The results also show that the features extracted during the MD walking condition were more than that of MS and MF among the important selected features. This suggests that in the MD walking condition (de-counting from 50), higher differences can be captured between fallers and non-fallers compared to the baseline and the secondary task of naming animals.

TABLE 3 – Feature Sets Selected by Relief-F and SBS

Feature Selection	**ML Algorithm**	**Feature Sets**
Relief-F	All	M_ST (MS), DC (MD),Diff_P (MD), Gender, STD_ST (MS), M_ST (MF), Diff_P (MF) M_ST (MD),STD_ST (MD)
SBS	KNN	M_ST (MS), DC (MD), Gender, SR_R (MF), STD_ST (MD), M_ST (MD), Diff_P (MD), US_L (MD)
SBS	SVM	Diff_P (MD), DC (MD), Gender, Diff_P (MF), M_ST (MS), SR_L (MD) STD_ST(MD), DTC_L (MD), PW_R (MD)
SBS	ANN	DC (MD), Gender, Diff_P (MD), US_L (MD), SR_R (MD), M_ST (MS) Diff_P (MF), M_ST(MF), STD_ST(MD)
SBS	Decision Trees	DC (MD), Gender, Diff_P (MD), M_ST (MD), STD_ST (MD), US_L (MF), Skx_R (MS), Diff_P (MF), Range_L (M), M_ST (MS)
SBS	Logistic Regression	M_ST (MS), Gender, PW_R (MD), M_ST(MD), STD_ST(MD), Diff_P (MD), Diff_P (MF), DC (MD), Skw_L (MF), DTC_L (MS), Range_R (MF)

The results shown in table 4 exhibit the performance of the classification models using the different feature sets. When using all the 43 features, ANN achieved the best performance with all the 43 features as inputs with 75.93% accuracy, 77.78% sensitivity, 74.07% specificity, and 75.00% precision.

The ANN model performance improved to 84.11% accuracy, 79.63% sensitivity, 88.68% specificity, and 87.76% precision with the feature reduction done using Relief-F. Furthermore, the SBS improved the performance to 89.81% accuracy, 90.74% sensitivity, 88.89% specificity, and 89.09% precision.

The results show that the ANN classification model performed the best out of the classification models explored in our study. This performance was improved by around 10% using the SBS feature selection techniques.

The statistical paired t-test for pairwise comparison was computed to test the differences between the different algorithms, and it was confirmed that there were statistically significant differences between the ANN model and the other classification models with 95% confidence.

TABLE 4 – Results of the Classification Models

Feature Set	Classification Model	Accuracy%±SD%	Sensitivity%±SD%	Specificity%±SD%	Precision%±SD%
All	KNN	63.89% ± 0.98%	61.11% ± 1.27%	66.67% ± 0.26%	64.71% ± 0.99%
Selected by Relief-F		70.37% ± 2.78%	66.60% ± 3.74%	74.07% ± 2.27%	72.00% ± 1.88%
Selected by SBS		84.266% ± 2.73 %	90.74% ± 2.95%	77.78% ± 2.08%	80.33% ± 3.11 %
All	SVM	65.74% ± 0.47%	61.11% ± 0.52%	70.37% ± 0.85%	67.35% ±0.37 %
Selected by Relief-F		73.15% ± 1.35%	66.67% ± 2.14%	79.63% ± 3.07%	76.605% ± 2.72%
Selected by SBS		85.19% ± 0.84%	83.33% ± 0.68%	87.04% ± 1.24%	86.54% ± 0.86%
All	**ANN**	75.93% ± 1.34%	77.78% ± 2.47%	74.07% ± 1.39%	75.00% ± 1.52%
Selected by Relief-F		84.11% ± 3.95%	79.63% ± 3.01%	88.68% ± 2.41%	87.76% ± 2.56%
Selected by SBS		**89.81**% ± **2.78**%	**90.74**% ± **2.14**%	**88.89**% ± **3.06**%	**89.09**% ± **2.67**%
All	Decision Tree	65.74% ± 0.88%	61.11% ± 1.24%	70.37% ± 0.52%	67.35% ± 1.81%
Selected by Relief-F		73.15% ± 3.27%	64.8% ±2.11%	81.48% ± 3.58%	77.78% ± 2.49%
Selected by SBS		83.33% ± 1.48%	81.48% ± 2.13%	85.19% ± 1.11%	84.62% ± 2.06%
All	Logistic Regression	64.81% ± 0.46%	61.11% ± 0.67%	68.52% ± 0.99%	66.00% ± 1.02%
Selected by Relief-F		72.22% ± 3.13%	65.81% ±3.38%	79.63% ± 2.47%	76.09% ± 2.64%
Selected by SBS		80.56% ± 2.32%	83.33% ± 2.14%	77.78% ± 2.68%	78.95% ± 2.04%

4 CONCLUSION

Results in this study show that the Degree of Cyclostationarity, gender, the mean and standard deviation of stride time of pressure walking signals are significant predictors for elderly fallers. We also showed that using feature selection methods improved the performance of the model and implementing hyperparameter tuning using grid search on each of the five classification models to obtain an optimized architecture for each classification model. Our analysis showed that the proposed model outperforms existing models by at least 10%. It is worth noting also that all performance measures of the selected model are above 88%, which is not the case for most existing models that increases one measure while reducing the others.

As a future perspective, we plan to explore additional physiological data and cyclostationary features as inputs to the classification models. We also want to try different approaches for feature selection since the SBS method is computationally expensive.

Références

[1] L. Z. Rubenstein, "Falls in older people : epidemiology, risk factors andstrategies for prevention", Age and ageing, vol.35, 2006.

[2] J. Howcroft, E. D. Lemaire, and J. Kofman, "Prospective elderly fallprediction by older-adult fall-risk modeling with feature selection", Biomedical Signal Processing and Control, vol. 43, pp. 320–328, 2018.

[3] K. Sabri, M. El Badaoui, F. Guillet, A. Belli, G. Millet, and J. B.Morin, "Cyclostationary modeling of ground reaction force signals", Signal Processing, vol. 90, no. 4, pp. 1146–1152, 2010.

[4] F. Zakaria, C. Toulouse, M. El Badaoui, C. Serviere, and M. Khalil,"Contribution of the cyclic correlation in gait analysis : Variation betweenfallers and non-fallers", in 2014 IEEE 16th International Conference one-Health Networking, Applications and Services (Healthcom), 2014, pp. 176–181.

[5] F. Zakaria, M. El-Badaoui, S. Maiz, F. Guillet, M. Khalil, K. Khalil,and M. Halimi, "Walking analysis : Empirical relation between kurtosis and degree of cyclostationarity", in 2013 2nd International Conferenceon Advances in Biomedical Engineering. IEEE, 2013, pp. 93–96.

[6] R. Brome, J. Nasreddine, F. Bonnardot, M. O. Diab, and M. El Badaoui, "Prediction of Elderly Falls Using the Degree of Cyclostationarity of Walk Pressure Signals". in 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). IEEE 2021, pp. 477-482.

[7] R. Brome, J. Nasreddine, F. Bonnardot, M. O. Diab, and M. El Badaoui, "Identifying Elderly Patients at Risk of Falling using Time-Domain and Cyclostationarity Related Features ", IJIE, vol. 13, no. 5, pp. 57–66, May 2021.

[8] Lee, Y. S., Moon, P. J. (2019). Analysis of Open-Source Hyperparameter Optimization Software Trends. International Journal of Advanced Culture Technology, 7(4), 56-62.

[9] F. Bonnardot, M. El Badaoui, R. Randall, J. Daniere, and F. Guillet,"Use of the acceleration signal of a gearbox in order to perform angularresampling (with limited speed fluctuation)", Mechanical Systems and Signal Processing, vol. 19, no. 4, pp. 766–785, 2005.

[10] G. Chandrashekar and F. Sahin,(2014). A survey on feature selection methods. Computers Electrical Engineering, 40(1), 16-28.