© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Identifying Elderly Patients at Risk of Falling using Time-Domain and Cyclostationarity Related Features

Reem Brome[1*], Jad Nasreddine[2], Frédéric Bonnardot[1], Mohamad O. Diab[2], Mohamed El Badaoui[1]

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

[2]Computer & Information Systems Department,
Rafik Hariri University, Meshref, LEBANON

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2021.13.05.007
Received 5 April 2021; Accepted 30 April 2021; Available online 31 July 2021

Abstract: Falls are a prevalent and severe health problem in the elderly community, leading to unfortunate and devastating consequences. Some falls can be prevented through interventions, proper management, and extra care. Therefore, studying and identifying elderly people with risk of falls is essential to minimize the falling risk and to minimize the severity of injuries that can occur from these falls. Besides, identifying at-risk patients can profoundly affect public health in a positive way. In this paper, we use classification techniques to identify at-risk patients using pressure signals of the innersoles of 520 elderly people. These people reported whether they had experienced previous falls or not. Two different types of feature sets were used as inputs to the classification models and were compared: The first feature set includes time-domain, physiological, and cyclostationary features, whereas the second includes a subset of those features chosen by Relief-F as the most important features. Our study showed that the use of features from different walking conditions and using Relief-F as a feature selection method significantly improved the model prediction accuracy, i.e. by 5.24% from the best previously existing model. The results also point out that the mean and standard deviation of the stride time, gender, the degree of cyclostationarity were the most important features to include in classification models for the identification of elderly people at risk of falling.

Keywords: classification models, cyclostationarity, degree of cyclostationarity, elderly people, fallers, machine learning, walk pressure signals

1. Introduction

The World Health Organization's (WHO) facts sheet (updated in January 2018) declared falls as the main cause of accidental injuries, which leads to deaths globally [1]. The WHO defines falling as an incident that result in an individual becoming at rest inadvertently on the ground or other lower levels. Unfortunately, fatality can occur due to injuries caused by falls. An estimated number of 646 thousand people worldwide die every year from accidental falls. Around 37.3 million fall cases demand medical assistance and attention yearly. The most notable number of fatal falls involved older adults with ages above 65 years old [1].

Rubenstein studied the common causes for falls in the elderly community [2]. His results revealed that some of these risk factors include, but are not limited to, being 80 years old or above, muscle defects, arthritis, depression, prior

*Corresponding author: reem.abou.marak.dit.brome@univ-st-etienne.fr

falls, the use of multiple medications, the use of an assistive device, impairments in stride movements, balance, cognition, and vision. Most inadvertent elderly falls occur due to a combination of various factors. He also mentioned that the significance of falls within the elderly community is not just limited to the point that the frequency of number of falls increases as age increases, but also that the severity of the injury is highest among the oldest subjects with a history of multiple prior falls. This leads to an increase in medical services and rehabilitation expenses. More importantly, it increases the possibility of disabilities and fatalities [2]. Therefore, reducing the risk of accidental 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 [2].

In this context, there is growing interest in the prediction of future elderly falls to help reduce its risk. Swanenburg, de Bruin, Uebelhart, Mulder [3] studied whether force plate variables can be employed to predict multiple fallers using statistical analysis. They noted that the amplitude of medial-lateral movements in single-task conditions was a significant independent predictor of elderly fallers, along with having a history of multiple falls [3]. Howcroft, Lemaire, Kofman, and McIlroy [4] found statistically significant differences in the static posturography measures between those that experienced accidental falls in the past and those who had not. In their study, they investigated eyes open, and eyes closed standing posturography with elderly adults and were able to identify differences and determine the measure cut-off scores for classifying prospective fallers, single-fallers, multi-fallers, and non-fallers [4].

Howcraft, Lemaire, and Kofman [5] also studied the prediction of elderly fallers using machine-learning algorithms where their best classification model achieved 65% accuracy and 59% sensitivity, using pressure-sensing-insole and left-shank-accelerometer as predictors [5]. Pressure sensing insoles have been also used throughout the literature to predict or analyze chronic medical conditions and diseases in the elderly community, such as Parkinson's disease [6], dementia [7], and accidental falls [8].

Properties of cyclostationarity in gait signals have been also used in modeling and analyzing human walk and ground reaction force (GRF) signals [9] [10] [11]. Sabri et al. [9] proposed an alternative framework for the study of GRF signals, based on cyclostationary characteristics rather than the traditional use of signal processing methods, which assume statistically stationary signal components. The proposed framework was able to model the periodicity of the signal statistics and showed improved results in demonstrating the development of runners' fatigue detection. Zakaria et al. examined and exploited the Cyclostationary (CS) properties and features such as the cyclic autocorrelation function. Their work demonstrated that there is a significant difference in the cyclic autocorrelation of fallers and non-fallers [10]. One indicator of cyclostationarity is the degree of cyclostationarity (DS) [12] [13]. The use of the average degree of cyclostationarity as a single feature in classification models proved to show promising results in previous work [14] with 68.43% accuracy using K-nearest neighbours method.

In this paper, we have investigated and compared the performances of different machine-learning classification models to identify older adults at risk of falling using different feature sets. The inputs used in this study include 14 features, 10 of which are classical time-domain features and the average degree of cyclostationarity explored in our previous work [14]. In addition, three new features are added due to their popularity in research studies related to gait analysis: the mean of the stride time, the standard deviation of the stride time, and gender. These 14 features were extracted from 3 different types of walking conditions: normal walking (MS), walking while fluently naming animals (MF), and walking while counting backward from 50 (MD). The performance of the classification model using all the 14 features with the three different types of walking was compared to that using a reduced number of predictors in each type of walking condition separately. The features were reduced using the Relief-F feature selection method [15]. The last feature set includes all the 14 features from all walking conditions with the gender being the same in all walking conditions to have a total of 40 features input to the classification model. The performance of that model was also compared to previous ones. Results show that utilizing Relief-F and applying classification methods on different sex independently improved system performance and reduced its complexity by selecting the features with the highest weights in prediction as inputs to the classification model.

The paper is organized as follows. In Section 2, we first describe the classification models built by first expexplain the experimental design and construction of the database. We then present the time-domain features, define the cyclostationary properties, and how the average degree of cyclostationarity can be obtained. In Section 3, we explain the methods used for Feature selection. In Section 4, the five different algorithms used for classification are also presented. In Section 5, we present and discuss the performance of the studied models. We finally summarize our findings, limitations, and prospects in the Section 6.

2. Model Description

In this study, five different classification algorithms are used with two different feature sets, in three different walking conditions, and with different types of features. This section details presents the processes of data collection, data preprocessing, feature extraction, and feature selection.

2.1 Experiment and Dataset Description

In this study, we used the dataset collected by the LPE (Laboratoire de Physiologie de l'Exercice) [16] and CHU (Centre Hospitalo-Universitaire) of Jean Monnet St-Etienne University [17]. The experiment was conducted on Elderly adults who were above the age of 75 years old. The participants wore a pair of innersole pressure sensors fitted inside their shoes. Each innersole contains two independent SMTEC electronic Foot switches system. These footswitches were fixed in specific positions to record the pressure signals from the heels and toes of the participants' left and right feet while walking [18]. These footswitches were connected to a portable data logger that the participants wore at their waist. In this setup, a pressure above 40 g/cm2 would activate the sensors and indicate the state of contact with the ground surface. For each step, the activation of the heel sensors establishes the first contact, whereas the last contact with the ground corresponds to when the toes sensor stops measuring pressure.

As mentioned by the manufacturer, the data is sampled from footswitches at a frequency of 100 Hz allowing a temporal resolution of 10 ms. The signals collected were processed using software designed specifically for the task by SMTEC software. The system measures four independent pressure signals: left heels, left toes, right heels, and right toes. 520 healthy elderly adults were recruited to partake in this experiment at the Hospital University of Saint Etienne [17]. They were all instructed to walk in the same straight 20 meters hallway while wearing these sensors. At first, they walked for a test trial. After that, each participant was asked to walk this same distance three times. The first time, which is the baseline in this study, they walked without performing secondary tasks, denoted MS walking condition. The second time, they walked the same distance again but while de-counting from the number 50, denoted MD walking condition. The third time, they walked while enumerating aloud as many animal names as they could remember, denoted MF walking condition. Some measures were taken while collecting the data to block other factors that could influence participants' walk. These measures included ensuring proper lighting, a quiet area, and the use of comfortable flat shoes.

Out of the 520 subjects, 302 were females and 217 were males. Only 54 subjects reported that they experienced previous falls, whereas the rest said they had not. As a first stage working with this largely unbalanced data, we included in our study the 54 fallers and we randomly chose 54 non-fallers to build classification models that can accurately classify elderly patients at risk of falling and those without the risk of falling.

2.2 Time Domain Features

The classical features extracted from the pressure signals were: mean, rise time, fall time, pulse width, overshoot, undershoot, duty cycle, slew rate, mid-cross, autocorrelation, standard deviation, band power, median, root mean square, range, P-welch, skewness, interquartile range, kurtosis, and 95 percentile of the signal distribution. The statistical significance of these features for elderly falling risk were tested using one-way repeated measures ANOVA [18]. The results showed that 10 features out of those had statistical significant differences between fallers and non-fallers, in at least 1 type of walking condition, were: pulse-width (right foot), undershoot (right and left feet), duty cycle (left foot), slew rate (right and left feet), range (right and left feet), and skewness (right and left feet).

2.3 Cyclostationary Analysis of Gait Signals

A cyclostationary process is a signal that has statistical properties varying cyclically with time. In studies related to human locomotion, the human walk can be described as a movement that consists of repeated replicated series of cyclic physical actions or strides. Analyzing the cyclostationary characteristics of the foot plantar pressure signals can introduce new features that signify the risk of falling in the elderly [10].

Using cyclostationarity properties requires a constant number of samples per stride. Therefore, it is important to pre-process the signal to compensate for the speed fluctuation. Compensating for the speed fluctuation is achieved by estimating this fluctuation as described in the work of Bonnardot et al. [20] and using interpolation to stretch the signal. There are two orders of cyclostationarity as demonstrated in the work of Spooner and Gardner [21]. A signal $S(t)$ is said to be cyclostationary of order 1 with cycle T if the expectation $\mu_{S(t)}$ of $S(t)$, is periodic with period T:

$$\mu_{S(t)}(t) = \mu_{S(t)}(t+T) = \mu_{S(t)} \qquad (1)$$

$\mu_{S(t)}$ represents the repetitive pattern in the signal. The residual signal $r(t)$ can be computed by removing $\mu_{S(t)}$ from the signal:

$$r(t) = S(t) - \mu_{S(t)} \qquad (2)$$

A signal is considered cyclostationary of order 2 if the autocorrelation $C_{S(t)}$ of the signal $S(t)$ is periodic with period T:

$$C_{S(t)}(t_2, t_1) = C_{S(t)}(t_1 + T, t_2 + T) \qquad (3)$$

where t is defined as:

$$t = (t_1 + t_2)/2 \qquad (4)$$

By defining τ as

$$\tau = t_2 - t_1 \quad (6)$$

We can write,

$$C_{S(t)}(t_2, t_1) = C_{S(t)}(t_1 + T, t_2 + T) = C_{S(t)}(t_2 - t_1) = C_S(t, \tau) \quad (5)$$

In the case of a cyclostationary signals of order 2 such as the pressure signals involved in this study, the instantaneous autocorrelation function is periodic, and therefore can be represented as a Fourier series as shown below [22]:

$$C_{S(t)}(t, \tau) = \sum_{\upsilon} CAF_S(\upsilon, \tau) e^{(-j2\pi\upsilon t)} \quad (7)$$

where, υ is the cyclic frequency that belongs to the set of cyclic frequencies such that, $\upsilon = k/T$ and $k \in \mathbb{Z}$. The Cyclic Autocorrelation Function (CAF) is defined as:

$$CAF_S(\upsilon, \tau) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} C_S(t, \tau) e^{(-j2\pi\upsilon t)} \quad (8)$$

In order to remove the influence of cyclostationarity at order 1 in cyclic autocorrelation, it is possible to compute the cyclic autocorrelation $CAF_R(\upsilon, \tau)$ of the residual signal $r(t)$, from Equation (2), instead of the autocorrelation $CAF_S(\upsilon, \tau)$ of the signal $S(t)$.
The CAF is a three-dimensional representation of the cyclostationarity of the pressure walking signal. It is possible to also obtain another 3-dimension representation called the spectral correlation $SCD_R(\upsilon, f)$ by taking the Fourier Transform of the CAF over τ to obtain a frequency f (figure 1). In this case, all the information are contained in the cyclic frequencies υ associated to the characteristics cycles of the signal. At other cyclic frequencies, the spectral correlation should have zero energy. This representation is complicated to use directly as input features for classification methods. Hence, the degree of cyclostationarity is used to summarize the cyclostationarity of the signal with less complexity. Zivanovic and Gardner [12] described the Degree of Cyclostationarity (DC) as a proper measure of the degree of non-stationarity for stochastic processes that exhibit cyclostationarity. The DC can be described as the distance to the closest stationary process exhibiting a comparable power spectral density. The cyclic frequency having non-zero energy on the cyclic statistics of order 2 is a parameter that proves the presence of a cyclostationary signal. The degree of cyclostationarity is defined by the following equation:

$$DC_S^{\upsilon} = \frac{\sum_f |SCD_R(\upsilon, f)|^2}{\sum_f |SCD_R(0, f)|^2} \quad (9)$$

$$DC_S = \sum_{\upsilon} DC_S^{\upsilon} \quad (10)$$

The pressure signals from the 4 sensors (toes and heels from right and left feet) are synchronized and transformed to obtain the degree of cyclostationarity DC_S^{υ}. An example of DC_S^{υ}such signal is shown in Figure 2. In this figure, the cyclic frequencies not associated to signal cycles are not zero but are negligible since this is an estimated representation. The summation DC_S can serve as a valuable single feature extracted from walk pressure signals for the identification of elderly fallers as shown in previous work [14] and is used in the classification models explained in the following sections.

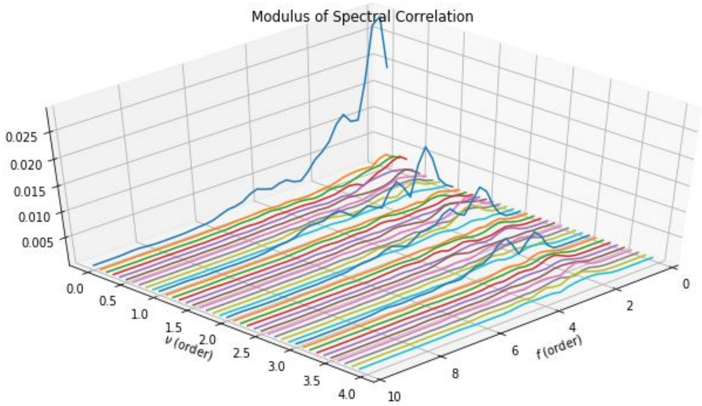

Fig. 1 - The spectral correlation of the average walk pressure signal during MS walking condition

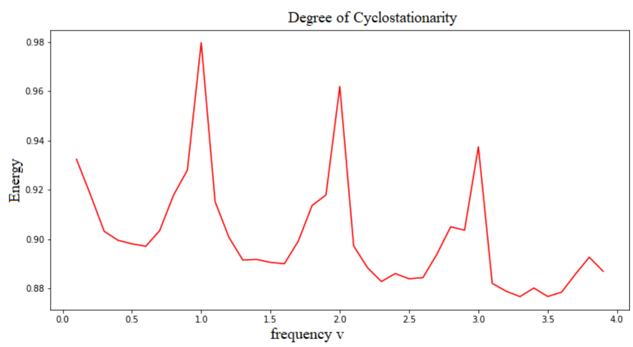

Fig. 2 - The degree of cyclostationarity of walk pressure signals during normal walking with frequency υ

3. Feature Selection

As a starting point, we considered the same features used in [14], which include the first 10 time-domain features, in addition to the DC feature (number 14 in Table 1). In [14], we showed using both statistical and machine learning methods that these features can be used to detect fallers in elderly people. After conducting further statistical and visual analysis on the dataset, we were able to identify two new categories of features: gender and stride time. Stride time is defined as the time elapsed between the first contact of two consecutive footsteps of the same foot and is expressed in milliseconds. In particular, we found the mean and standard deviation of the stride time are significant features to detect fallers and non-fallers. Therefore, we included these three features in our model.

Then, we constructed a second set of features using Relief-F method, which ranks the features according to their importance. The number of ranked best predictors were selected based on the number that yielded the highest classification accuracy in each set of features associated with each type of walking condition.

Relief-F first sets all predictor weights to 0. The algorithm penalizes the features that give different values to neighbors of the same class and rewards the features that provide different values to neighbors of different classes. The algorithm iteratively selects a random observation, finds the k-nearest observations to that observation for each class, and updates all the weights for the features. If a feature value difference is observed in a neighboring instance pair with the same class, the feature score decreases. On the other hand, if a feature value difference is observed in a neighboring instance pair with different class values, the feature score increases [15].

Table 1 - Features used in the classification models

Feature Reference Number	Feature Abbreviation	Feature Description
1	PW_R	Pulse Width of Right Foot
2	US_R	Undershoot of Right Foot
3	US_L	Undershoot of Left Foot
4	DTC_L	Duty Cycle of Left Foot
5	SR_R	Slew Rate of Right Foot
6	SR_L	Slew Rate of Left Foot
7	Range_R	Range of Right Foot
8	Range_L	Range of Left Foot
9	Skw_R	Slew Rate of Right Foot
10	Skw_L	Slew Rate of Left Foot
11	Gender	Male or Female
12	M_ST	Mean Stride Time
13	STD_ST	Standard Deviation of Stride Time
14	DC	Degree of Cyclostationarity

4. Classification Models

In this study, we used five classification methods: K-Nearest Neighbors [23], Support Vectors Machines [24] with polynomial kernels of degree 3 (as it yielded the highest average accuracies among the linear, and polynomials of kernels: 1, 2, and 4), Artificial Neural Networks [25] with 10 nodes in a single hidden layer (as this architecture for one hidden layer scored the highest in the average accuracies with the lowest number of nodes possible), Decision Trees [26], and Logistic Regression [27].

5. Results and Discussion

A 100 times 10 folds cross-validation was applied on all classifier models. The results obtained are shown in Table 2 for the MS normal walking condition, Table 3 for the MF condition, and Table 4 for the MD walking condition. Finally, Table 5 shows the results for all the features in all types of walking conditions. The first feature set includes all the 14 features mentioned in Table 1. The second feature set were selected using Relief-F. The number of best predictors used in the models was chosen based on the combination, leading to the highest classification accuracy, as shown in figures 3, 4, 5, and 6.

Table 2 - Results of the classification models for the MS walking condition

Feature Set	Model	Accuracy%	Sensitivity%	Specificity%	Precision%
All Features (14)	**KNN**	66.72%± 2.55%	62.96%± 1.23%	70.37± 7.96%	68%± 3.85%

Selected Features (12, 3, 13, 6, 9)		70.43%± 1.65%	68.52%± 4.69%	72.22%± 8.74%	71.15%± 3.56%
All Features (14)	**SVM**	69.40%± 5.55%	74.07%± 7.21%	64.81%± 4.14%	67.8%± 5.73%
Selected Features (12, 3, 13, 6, 9)		71.32%± 2.53%	77.78%± 5.74%	64.81%±6.39%	68.85%± 8.06%
All Features (14)	**ANN**	75.92% ± 5.41%	76.92% ± 1.35%	75.00%± 4.21%	74.07%± 6.42%
Selected Features (12, 3, 13, 6, 9)		**78.70%± 4.57%**	**78.18%± 4.29%**	**79.25%± 6.35%**	**79.63%± 7.58%**
All Features (14)	**Decision Tree**	63.12%± 4.37%	68.52%± 5.84%	57.41%±7.53%	61.67%± 4.47%
Selected Features (12, 3, 13, 6, 9)		62.35%± 4.02%	59.26%± 3.78%	64.81%± 7.58%	62.75%± 5.71%
All Features (14)	**Logistic Regression**	68.54%± 2.12%	74.07%± 5.78%	62.96%± 7.14%	66.67%± 5.98%
Selected Features (12, 3, 13, 6, 9)		69.44%± 5.01%	75.93%± 2.30%	62.96%± 7.84%	67.21%± 6.33%

Table 3 - Results of the classification models for the MF walking condition

Feature Set	Model	Accuracy%	Sensitivity%	Specificity%	Precision%
All Features (14)	**KNN**	68.72% ± 3.46%	60.04% ± 5.63%	76.17% ± 5.79%	71.32% ±7.80%
Selected Features (12, 13, 14, 3, 6, 11, 4, 9)		72.00% ± 2.28%	65.39% ± 4.87%	78.02% ± 3.67%	75.24% ± 5.35%
All Features (14)	**SVM**	67.92% ± 5.23%	69.02% ± 6.16%	67.91% ± 5.13%	67.56% ± 8.49%
Selected Features (12, 13, 14, 3, 6, 11, 4, 9)		72.37% ± 4.05%	80.55% ±2.57%	63.83% ± 6.00%	69.68% ± 7.08%
All Features (14)	**ANN**	71.86% ± 2.77%	75.63% ± 4.69%	70.06% ± 5.84%	64.96%± 5.63%
Selected Features (12, 13, 14, 3, 6, 11, 4, 9)		**75.37% ± 6.41%**	**74.37% ± 6.72%**	**76.54% ± 7.08%**	**76.95% ± 6.57%**
All Features (14)	**Decision Tree**	63.21% ± 5.32%	58.62%± 7.63%	66.06% ± 7.65%	63.05% ± 6.84%
Selected Features (12, 13, 14, 3, 6, 11, 4, 9)		68.64% ± 6.12%	76.37% ± 5.64%	63.07% ± 5.32%	67.33% ± 6.97%
All Features (14)	**Logistic Regression**	56.58% ± 3.04%	60.32% ± 2.63%	55.56% ± 8.32%	60.51% ± 3.19%
Selected Features (12, 13, 14, 3, 6, 11, 4, 9)		68.37% ± 3.64%	74.98% ± 5.63%	62.33% ± 4.97%	66.36% ± 7.63%

Table 4 - Results of the classification models for the MD walking condition

Feature Set	Model	Accuracy%	Sensitivity%	Specificity%	Precision%
All Features (14)	**KNN**	69.9% ± 6.19%	60.07% ± 7.01%	79.94% ± 6.88%	75.00% ± 7.24%
Selected Features (3, 14, 12, 11, 5, 13)		70.9% ± 4.22%	63.52% ± 4.68%	78.81% ± 5.64%	74.01% ± 4.46%
All Features (14)	**SVM**	67.98 ± 5.00%	67.24% ± 2.36%	69.52% ± 4.59%	68.39%± 5.63%
Selected Features (3, 14, 12, 11, 5, 13)		73.57% ± 6.18%	72.94% ± 6.49%	73.03% ± 4.81%	72.84% ± 5.39%
All Features (14)	**ANN**	75.05% ± 5.61%	72.67% ± 6.74%	73.12% ± 2.94%	75.99%± 3.78%

Selected Features (3, 14, 12, 11, 5, 13)		**79.93% ± 2.56%**	**77.81% ± 5.72%**	**77.78%± 5.28%**	**78.43%± 5.34%**
All Features (14)	**Decision Tree**	70.26% ± 2.37%	73.11%± 3.61%	65.64%± 4.88%	68.20%± 4.67%
Selected Features (3, 14, 12, 11, 5, 13)		71.31% ± 5.41%	71.33% ± 4.85%	71.36% ± 5.23%	72.66% ±5.23%
All Features (14)	**Logistic Regression**	67.17% ± 3.21%	67.81% ± 5.74%	67.24% ± 5.99%	67.83% ± 6.41%
Selected Features (3, 14, 12, 11, 5, 13)		68.63% ± 6.27%	70.35% ± 2.91%	69.83% ± 6.31%	69.11% ± 5.44%

Table 5 - Results of the classification models for all the walking conditions

Feature Set	Model	Accuracy%	Sensitivity%	Specificity%	Precision%
All Features (40)	**KNN**	63.45% ± 5.12%	55.23% ± 5.61%	75.62% ± 7.59%	70.56% ± 6.14%
Selected Features (12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		68.53% ± 1.92%	60.32% ± 4.58%	70.55% ± 6.33%	74.54% ± 6.81%
All Features (40)	**SVM**	64.82% ± 3.67%	68.73% ± 4.38%	62.75% ± 5.64	64.26% ± 5.32%
Selected Features (12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		66.59% ± 3.88%	68.96% ± 2.37%	62.40% ± 5.07%	64.78% ± 5.48%
All Features (40)	**ANN**	75.51% ± 2.74%	70.62% ± 2.99%	73.83% ± 5.20%	76.82% ± 4.17%
Selected Features (12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		**81.16% ± 2.87%**	**79.63% ± 5.63%**	**78.23% ± 5.11%**	**79.81% ± 4.82%**
All Features (40)	**Decision Tree**	63.47% ± 4.27%	63.98% ± 3.75%	65.32% ± 5.84%	63.71% ± 4.27%
Selected Features (12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		72.69% ± 2.86%	71.41% ± 3.78%	73.47% ± 5.21%	73.20% ± 3.68%
All Features (40)	**Logistic Regression**	64.50% ± 4.87%	68.87% ± 4.34%	59.89% ± 3.33%	63.11% ± 3.96%
Selected Features (12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF)		70.71% ± 3.04%	73.28% ± 3.85%	68.21% ± 4.25%	68.57% ± 5.42%

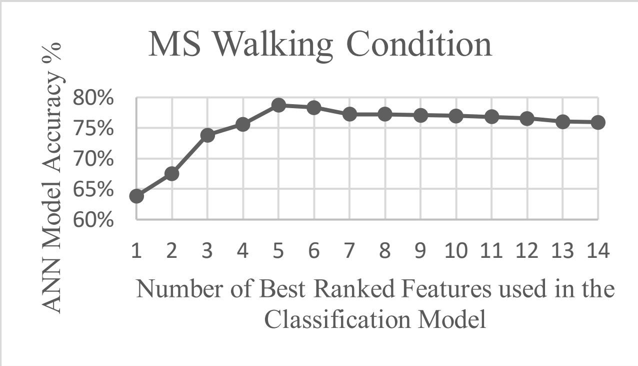

Fig. 3 - Number of best-ranked features in MS

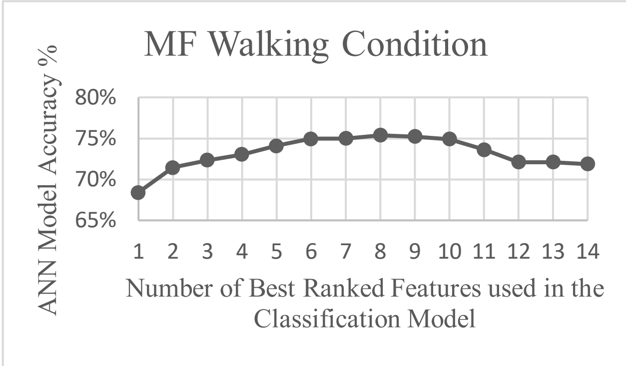

Fig. 4 - Number of best-ranked features in MF

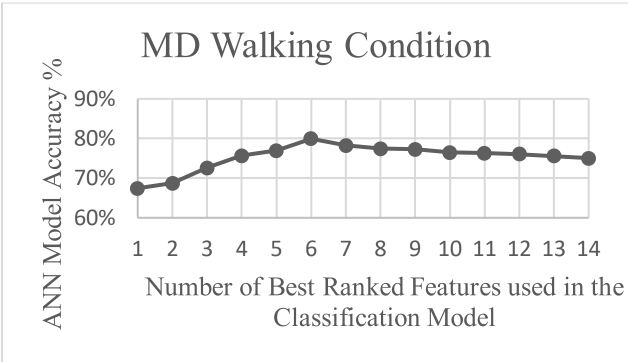

Fig. 5 - Number of best-ranked features in MD

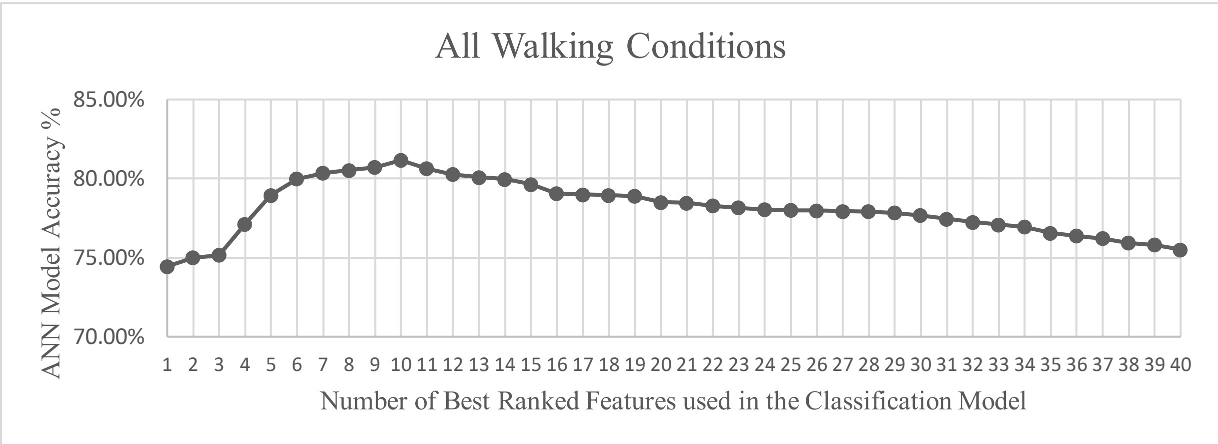

Fig. 6 - Number of best-ranked features in all walking conditions

Table 2 shows the results in the MS walking condition. The best performance with all the 14 features as inputs was 75.9% accuracy, 76.92% sensitivity, 75.00% specificity, and 74.07% precision using ANN as a classifier. This performance was improved using Relief-F where, as shown in figure 3, performed best with the first 5 best predictors for this model (12, 3, 13, 6, 9). With feature selection, the ANN model performance improved to 78.7% accuracy, 78.18% sensitivity, 79.25% specificity, and 79.63% precision.

Table 3 shows the results in the MF walking condition. The best performance with all the 14 features as inputs was 71.86% accuracy, 75.63% sensitivity, 70.06% specificity, and 64.96% precision also using ANN as a classifier. This performance was improved using the best 8 predictors for this model, as shown in figure 4 (which were: 12, 13, 14, 3, 6, 11, 14, 9). With feature selection, the ANN model performance improved to 75.37% accuracy, 74.37% sensitivity, 76.54% specificity, and 76.95% precision.

Table 4 shows the results in the MD walking condition. The best performance with all the 14 features as inputs was 75.05% accuracy, 72.67% sensitivity, 73.12% specificity, and 75.99% precision, also using ANN as a classifier. This performance was improved using Relief-F, which selected the best 6 predictors for this model, as shown in figure 5. (These features were: 3, 14, 12, 11, 5, 6). With feature selection, the ANN model performance improved to 79.93% accuracy, 77.81% sensitivity, 77.78% specificity, and 78.43% precision.

The results shown in Table 5 show the performance of the model using all the predictors from the 3 walking conditions combined. As shown in figure 6, the model accuracy scored the highest with the first 10 best-ranked features: 12_MS, 14_MD, 11, 13_MS, 12_MF, 12_MD, 13_MD, 3_MS, 13_MF, 14_MF. The best performance with all the 40 features as inputs was 75.51% accuracy, 70.62% sensitivity, 73.83% specificity, and 76.82% precision, also using ANN as a classifier. The ANN model performance improved to 81.16% accuracy, 79.63% sensitivity, 78.23% specificity, and 79.81% precision with the feature reduction done using Relief –F.

The results show that the ANN classification model performed the best out of the classification models explored in our study using the features chosen as inputs for the models. This performance was generally improved using the Relief-F feature reduction method. Furthermore, the MD walking condition helped achieving the highest accuracies compared to the MS and MF conditions. This points out that in the MD walking condition, which is walking while de-counting from 50, there are higher differences that can be captured between faller and non-fallers compared to the baseline and the secondary task of naming animals.

It is noticeable that the common most important features in prediction of elderly people at risk of falling were: The mean stride time (M_ST), standard deviation stride time (STD_ST), and Undershoot of left foot (US_L) in all the three walking conditions. The Degree of Cyclostationarity (DC) and gender were important features in the MF and MD walking conditions.

Using all the features from all walking conditions improved system accuracy after applying Relief-F and achieved the best performance in terms of accuracy with the ANN machine-learning algorithm. 10 features were selected out of the 40 features: one of them is gender, and 3 features from each type of walking condition. This points out that combining the 3 types of walking conditions in one classification model helped in achieving better performance than considering each type of walk condition by itself.

The statistical paired t-test for pairwise comparison was computed and it was confirmed that there were statistical significant differences between the ANN model in each type of walking condition and the other classification models with the same type of walking condition. In addition, the statistical independent t-test was applied to confirm statistical significant differences between the ANN models in the MS, MF, MD, and the case involving all walking conditions.

6. Conclusion

In our previous work, we showed that including the Degree of Cyclostationarity could improve model performance in the classification of elderly fallers and non-fallers. In this study, we included three additional features: Mean of the stride time, Standard Deviation of the stride time, and Gender. In addition, we used Relief-F method for feature selection. We compared the results of different classification models in terms of feature, type of walking condition, and methods of classification. The use of the Relief-F method reduced the features and selected the best out of them. This step generally improved and optimized the classification model. Results show that some of the best features for prediction of elderly people at risk of falling were the stride time in all walking conditions, and the degree of cyclostationariy and gender in the MF and MD walking condition. Combining the features of all types of walking conditions in the ANN classification model along with using Relief-F feature selection method lead to the best performance in terms of accuracy.

As a future perspective, we intend to include additional physiological data and cyclostationary features as inputs to the classification models and compare their outcomes.

Acknowledgements

We gratefully thank Prof. Frédéric Roche, Laboratoire de Physiologie de l'Exercice [16] and Centre Hospitalo-Universitaire [17] of Jean Monnet St-Etienne University for providing the database used in this study.

References

[1] W. H. Organization, W. H. O. Ageing, and L. C. Unit, WHO global report on falls prevention in older age. World Health Organization, 2008.
[2] Rubenstein, L. Z. (2006). Falls in older people: epidemiology, risk factors and strategies for prevention. Age and ageing, 35(suppl_2), ii37-ii41..
[3] Swanenburg, J., de Bruin, E. D., Uebelhart, D., & Mulder, T. (2010). Falls prediction in elderly people: a 1-year prospective study. Gait & posture, 31(3), 317-321.
[4] Howcroft, J., Lemaire, E. D., Kofman, J., & McIlroy, W. E. (2017). Elderly fall risk prediction using static posturography. PLoS one, 12(2), e0172398
[5] Howcroft, J., Lemaire, E. D., & Kofman, J. (2018). Prospective elderly fall prediction by older-adult fall-risk modeling with feature selection. Biomedical Signal Processing and Control, 43, 320-328.

[6] Alam, M. N., Garg, A., Munia, T. T. K., Fazel-Rezai, R., & Tavakolian, K. (2017). Vertical ground reaction force marker for Parkinson's disease. PLoS One, 12(5), e0175951.
[7] Bharucha, A. J., Anand, V., Forlizzi, J., Dew, M. A., Reynolds III, C. F., Stevens, S., & Wactlar, H. (2009). Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges. The American journal of geriatric psychiatry, 17(2), 88-104.
[8] Takahashi, T., Ishida, K., Hirose, D., Nagano, Y., Okumiya, K., Nishinaga, M. & Yamamoto, H. (2004). Vertical ground reaction force shape is associated with gait parameters, timed up and go, and functional reach in elderly females. Journal of rehabilitation medicine, 36(1), 42-45.
[9] Sabri, K., El Badaoui, M., Guillet, F., Belli, A., Millet, G., & Morin, J. B. (2010). Cyclostationary modeling of ground reaction force signals. Signal Processing, 90(4), 1146-1152.
[10] Zakaria, F. A., Toulouse, C. V., El Badaoui, M., Serviere, C., & Khalil, M. (2014, October). Contribution of the cyclic correlation in gait analysis: Variation between fallers and non-fallers. In 2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom) (pp. 176-181). IEEE.
[11] Toulouse, C. V., El Badaoui, M., & Serviere, C. (2015, September). Gait analysis and falls estimation using the slope's variation of synchronous statistics. In 2015 International Conference on Advances in Biomedical Engineering (ICABME) (pp. 293-296). IEEE.
[12] Živanović, G. D., & Gardner, W. A. (1991). Degrees of cyclostationarity and their application to signal detection and estimation. Signal processing, 22(3), 287-297.
[13] Zakaria, F. A., El-Badaoui, M., Maiz, S., Guillet, F., Khalil, M., Khalil, K., & Halimi, M. (2013, September). Walking analysis: Empirical relation between kurtosis and degree of cyclostationarity. In 2013 2nd International Conference on Advances in Biomedical Engineering (pp. 93-96). IEEE.
[14] Brome, R., Nasreddine, J., Bonnardot, F., Diab, M. O., & El Badaoui, M. (2021, March). Prediction of Elderly Falls Using the Degree of Cyclostationarity of Walk Pressure Signals. In 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES) (pp. 477-482). IEEE.
[15] Robnik-Šikonja, M., & Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. Machine learning, 53(1), 23-69.
[16] Laboratoire de Physiologie de l'Exercice EA4338, Université Jean Monnet, Saint-Etienne F-42023, Université de Lyon, France.
[17] CHU Centre Hospitalier Universitaire de Saint- ´ Etienne - Loire - Accueil. [Online]. Available: http://www.chu-st-etienne.fr/. [Accessed: 14- Oct-2020].
[18] Zakaria, F. (2015). Human locomotion analysis: exploitation of cyclostationarity properties of signals (Doctoral dissertation, Saint-Etienne).
[19] Gaddis, M. L. (1998). Statistical methodology: IV. Analysis of variance, analysis of co variance, and multivariate analysis of variance. Academic emergency medicine, 5(3), 258-265.
[20] Bonnardot, F., El Badaoui, M., Randall, R. B., Daniere, J., & Guillet, F. (2005). Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). Mechanical Systems and Signal Processing, 19(4), 766-785.
[21] Spooner, C. M., & Gardner, W. A. (1994). The cumulant theory of cyclostationary time-series. II. Development and applications. IEEE Transactions on Signal Processing, 42(12), 3409-3429.
[22] Bonnardot, F., C. C., and R. L. (2018) Whole Body Vibrations Physical and Biological Effects on the Human Body.CRC Press, ch. Signal Processing Applied to the Analysis of Vibrations Transmitted in the Human Body.
[23] Tartakovsky, D. M., Wohlberg, B., & Guadagnini, A. (2007). Nearest ‐ neighbor classification for facies delineation. Water resources research, 43(7).
[24] Muhammad, I., & Yan, Z. (2015). Supervised Machine Learning Approaches: A Survey. Ictact Journal on Soft Computing, 5(3).
[25] Sarle, W. S. (1994). Neural networks and statistical models.
[26] De Ville, B. (2013) "Decision trees", Wiley Interdisciplinary Reviews: Computational Statistics, vol. 5, no. 6, pp. 448–455.
[27] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons.