

An approach to Gait Analysis from Human-Rollator Interaction: The *i*-Walker

Atia CORTÉS ^{a,1}, Maribel OJEDA ^b, Javier BÉJAR ^a and Antonio B. MARTÍNEZ ^c

^a *Knowledge Engineering and Machine Learning Group (KEMLG), UPC*

^b *Instituto Tecnológico de Estudios Superiores de Monterrey, Puebla (México)*

^c *Grup de Recerca en Robòtica Intel·ligent i Sistemes (GRINS), UPC*

Abstract. Falls are one of the main causes of injury and injury-related deaths in the older demographic sector. During the last decade, several solutions have been proposed for fall detection or prevention, but also to assist old adults in their mobility. The main concern is to provide means to improve their quality of life, allowing them to live in community as long as possible, but also to help maintaining public health systems as sustainable as possible. In this paper, we compare two different methodologies that aim to define walking patterns in 42 adults aged 22 to 94 years that participated in a pilot study. Two methodologies are compared: one focused on the whole exercise performance, while the other tries to identify strides and categorize them. Both proposals use a technique based on the bag-of-words approach.

Keywords. gait analysis, time series clustering, assistive technologies, bag-of-words, health informatics

1. Introduction

The ability to walk normally is related to several bio-mechanical components involved in the gait cycle (also known as *stride*), including (i) free mobility of joints, particularly in the legs; (ii) coordination of muscle action in terms of timing and intensity; (iii) normal sensory input, such as vision and vestibular system [1]. Thus, gait requires input from the brain, spinal cord, peripheral nerves, muscular power and joint and cardiovascular health. Because all of these systems are required to coordinate gait, the individual's walking speed is an indicator of the health of many physiological systems [2]. The relation between gait and cognition has been widely analysed from the medical point of view (see [3,4]). As people age, they tend to slow their gait speed, and their balance is also affected. Moreover, elderly people suffer different kinds of cognitive decline, which also accentuate gait disorders and its consequences [5]. Also, current concepts in disablement emphasise the importance of identifying mobility impairments in ageing humans to enable timely intervention and, ultimately, prevent disability. One of the most common and dramatic consequences of gait disorders is falling. Over a third of the population aged 65+ years fall every year (50% for adults aged 80+ years). As a consequence, 4-15% of falls cause significant injuries, while 23-40% of injury-related deaths in older adults are

¹Corresponding Author: Atia Cortés, Universitat Politècnica de Catalunya, Barcelona, Spain; E-mail: acortes@cs.upc.edu.

due to a fall [7]. Assistive devices, such as walkers, wheelchairs or canes, aim to provide physical support to challenged individuals and reduce the fear of falling, allowing them to maintain their physical activity and autonomy as long as possible.

Human locomotion has been studied for decades, although the perspective, along with the tools used for measurement, have evolved to these days. Gait analysis is a systematic technique for recognising negative deviations in the gait pattern and determining their reason and effects [8]. Several conditions might affect the ability to walk, and it affects mainly to people when ageing. It is essential that clinicians regularly assess gait in older adults to diagnose and plan optimal treatments for each situation. Although it is not possible to provide a general description of *gait* without including all the singularities given in each pathology, it is well accepted that a *normal gait* involves the locomotive action of the two legs, alternately to provide both support and propulsion, having always at least one foot in contact with the ground [9]. The walking activity is composed of the stance and swing phases: the faster we walk, the shorter the stance phase will be. A *gait cycle*, or stride, is thus defined as the time interval between two successive occurrences of a repetitive event (*e.g.* the right leg initial contact with the ground).

With the introduction of wearables and smartphones, it is nowadays easier to collect continuous and ubiquitous biometric data, in addition to lifestyle routines. Smartphones already provide personal health information, such as the number of steps or kilometres walked during the day. This measurement is done with inertial sensors (*i.e.* accelerometers, gyroscopes), which are nowadays found behind many commercial solutions of human motion tracking, but have also become very popular in the research field of gait analysis due to their reduced price and size, which allows different localisation possibilities in the human body (*e.g.* wrist, waist, ankle, foot insoles). Accelerometers, along with other biometric sensors, have been widely used to develop new quantitative approaches that provide empirical, objective results. Smart walkers, such as the *i-Walker*, offer the opportunity to embed sensors to their framework, providing means of monitoring and following-up elders activity while walking.

In this paper, the *i-Walker* [12] has been used as a measuring tool in a pilot study with 42 participants aged 22 to 94 years old performing a Three-Minutes Walking Test (3mWT) in an indoor corridor. The aim of the study was two-fold: first, to validate a step identification method that was already presented in [10] and a clustering approach using the bag-of-X technique with those obtained steps; second, to test another gait analysis proposal that has been described in [11]. The general objective is to find a methodology that allows to form groups of individuals by their walking characteristics, identifying those at high risk of falling. It is expected that the *i-Walker* could assist experts with the diagnose or rehabilitation treatment of old adults as part of the decision support system.

The structure of the paper is as follows: Section 2 describes the data collection process used in this study; Section 3 describes both methodologies and their differences; Section 4 shows the obtained results and some conclusions are given in Section 5.

2. Pilot Study

The work presented here was carried out at the Centre de Vida Independent (CVI), a care centre in Barcelona. The study involved 42 participants aged 22 to 94 years old (34 women and 8 men). From these, 25 were 65+ years old and 12 of those presented dif-

ferent health conditions, all related to cardiological pathologies, being the most frequent hypertension and vascular dementia; most of them had also suffered more than one fall during the last year. These characteristics were only present on adults older than 80 years that were permanently living at CVI; the rest of participants were relatives or workers of the centre. Participants have been categorized by (i) age (65- years (Y), 65-80 (M), 85+(O)); (ii) gender (Female and Male) and (iii) risk of falling based on the Tinetti scale [18] (Low, Medium or High).

Timed walking tests are commonly used in clinical assessments to evaluate the physical condition of individuals, especially elder adults or people rehabilitating from an injury. The studied outcome is usually the gait velocity during the exercise, but thanks to the inclusion of sensors or video recordings, it is also possible to extract spatio-temporal gait characteristics. Moreover, the relation between gait velocity and cognition has been widely studied [4,8]. The objective of this trial was to study the interaction of the participants with the *i-Walker* through the force sensors embedded in the handlers while walking. It is expected to find differences by age, gender or health condition. The main instructions given to the participants to perform the test were:

- Walk for three minutes along the corridor.
- Do not drop the handlers while performing the exercise.
- Turn when reaching the end of the hallway to keep the walking trajectory.
- If the time finishes and the person is at the middle of the corridor keep walking till returning to the starting point.

The measurement tool used in this test was the *i-Walker*, which records the data generated by each of its sensors onboard at 10Hz and stores it in *.csv* files. In this work, we focus on (i) the odometry data extracted from the rear wheel encoders to capture the estimated position, and therefore the resulting trajectory of the whole exercise; (ii) the gyroscope to detect the end of a straight line and beginning of the next one and discard the turning phases; and (iii) the pushing forces exerted by the user on the handlebars while walking. Each participant performed between 2 and 6 rounds, depending on their gait velocity. After a data cleaning process, our dataset is composed by 153 exercises (or rounds) performed by 42 individuals.

3. Methods

The main differences among all the studies found in literature about sensor-based gait analysis are:

- Components of the study: age population, number of participants, test performed
- Measurement tool: type of sensor(s) or data collected, sample time
- Gait recognition algorithm: how the data is processed to proceed with the objective of the analysis

In this paper we compare the results of two analysis that were carried out with the data collected at the CVI pilot. The first analysis takes each exercise (each straight line) as a whole and uses the three forces of each handlebar as inputs. A Bag-Of-SFA-Symbols (BOSS) model [6] was used for data representation and indexing by transforming time series (the straight lines) into a vocabulary based on the behaviour of data in the fre-

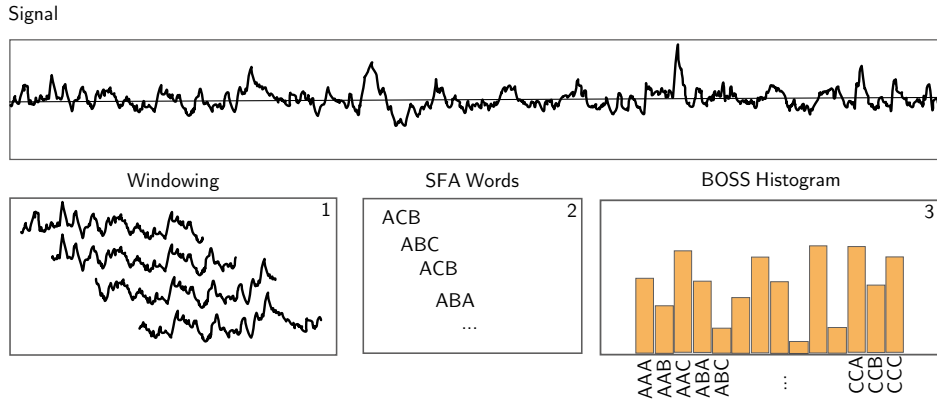


Figure 1. BOSS Model is used for indexing and representation and transforms the time series into BOSS histograms

quency domain (SFA stands for Symbolic Fourier Approximation, see Figure 1). Briefly, a set of overlapping windows are extracted from each signal. The FFT of each window is computed, and an equal frequency discretization is applied to the M first Fourier coefficients of all windows independently to obtain M sets of I discrete intervals. A word is assigned to each window consisting in the concatenation of the discretized values of their Fourier coefficients, these are the SFA symbols. Each signal is represented as the histogram of the words of their windows. A similarity is defined among the signals using the cosine similarity among the histograms. The data is then transformed by applying Spectral Embedding to their similarity matrix and then clustered with a Bayesian Dirichlet Process Gaussian Mixture Model. A more detailed description of the methodology and the obtained results are given in [11]. In this previous work, exercises were considered as a whole, discarding the curved zones and grouping the final straights into a single one. In this paper, each round is included as a single exercise in order to match it with the following analysis.

The second methodology presented in this paper studies the behaviour of the pushing forces applied to the *i-Walker* at each stride. The most critical step in gait identification is the segmentation of the raw data provided by the sensors into steps or strides that will be then analysed. Gait detection can be obtained with different techniques, such as frequency-based algorithms, gait cycle identification, local maxima among others. Research on gait analysis has lately been focused on pattern recognition and classification using different machine learning techniques [13]. In [10], we already introduced an approach to gait identification in straight lines using the 10 Meter Walk Test. Although results were promising, the short duration of the exercise did not allow a proper characterization of the participant's behaviour while walking. This is the reason to expand the timed-walking assessment to three minutes. Our gait identification approach used the combination of the pushing forces exerted by the individual while walking to detect local maxima peaks:

$$F_x dif f = rhfx - lhfx$$

where $rhfx$ corresponds to the right hand pushing force and $lhfx$ is the left hand pushing force (see Figure 2 top). The process has been automatized to capture strides of different

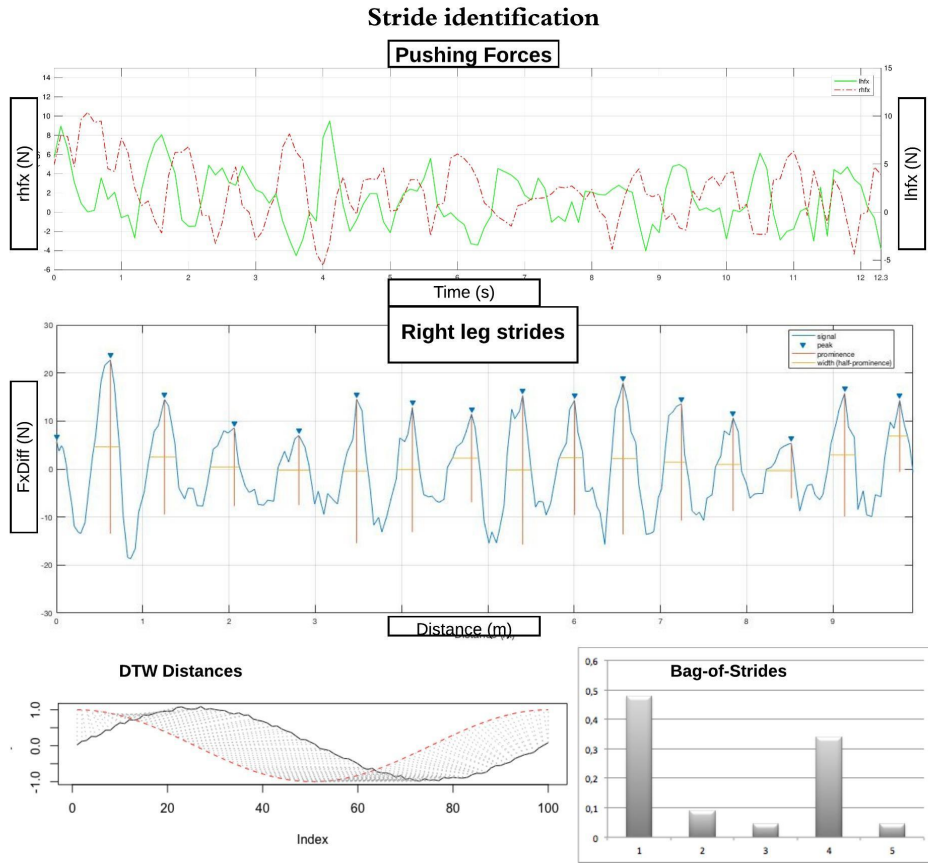


Figure 2. Stride identification using local maxima (top). DTW distances between strides are used as k-Medoids input to obtain strides clusters (bottom left). Bag-of-Strides for each exercise are used cluster exercises using KL-Divergence.

lengths based on the average gait velocity. Each stride is then collected as a separated time series with $F_x diff$ pushing forces applied. As a result, we obtain a new dataset of 4591 strides of different sizes. This dataset will be used to calculate the affinity matrix based on the shape of the pushing forces by using the Dynamic Time Warping distance (DTW [14], see Figure 2 bottom left). DTW finds the optimal alignment between two time series, regardless of their respective length. In fact, DTW has already been used in the context of gait analysis (see [15,16]). A DTW distance is calculated for each pair of strides from the strides dataset, this results in a $M \times M$ distance matrix, where M is the number of objects in the time series dataset.

The resulting DTW distance matrix is then used as input of a k -Medoids algorithm, instead of using the original strides dataset. This allows to categorize the strides into clusters using values of $k=\{2..6\}$. Each exercise can now be represented by the distribution of its strides in the different obtained k groups. As a result, a new dataset is created,

with as many objects as exercises had the original dataset, where each object represents a pair $\{user, exercise\}$ and the representation of the strides extracted from that exercise in terms of clusters, *i.e.* the proportion of strides of an exercise that falls in each stride cluster. In other terms, we obtain a histogram for each pair $\{user, exercise\}$ representing the frequency of each stride cluster by exercise. A set of histograms $H = \{h_1, h_2, \dots, h_n\}$ is generated, where each h_k is composed by a number of bins which sum up to one.

Clustering histograms has become popular thanks to the bag-of-words categorisation method. In general, this technique relies on identifying relevant key-words and analysing their frequency of appearance. In this paper, the stride vocabulary will be used to characterise types of exercises according to the occurrences of each type of stride. In this case, a bag-of-strides is associated to each exercise (see Figure 2 bottom right). The concept of *bag-of-steps* was introduced in [17], where authors used it to predict the rehabilitation length and discharge date of a patient using insole force sensors. Other than the final application, these two methodologies differ on the vocabulary generated (strides *vs* steps) and the tools used to collect them.

This technique also allows to extract the spatio-temporal characteristics of each stride and of the related exercise: stride length (cm) and time (s), the total distance and duration of the exercise, the number of strides, the average walking speed (m/s) and the cadence (strides/min). The last step is to apply a second k -Medoid clustering to our data, using the bags-of-strides histograms as input and the Kullback-Leibler Divergence as similarity measure [14].

One of the main challenges when using unsupervised learning techniques is to find the right number of clusters. In our case we have several combinations of the two clustering processes (first, applied to strides using DTW distances and then to exercises using KL-Divergence). A clustering stability analysis was applied to determine the final combinations of types of strides and exercises.

4. Results

After the stability analysis, several combinations of types of strides and exercises were given. However, the BOSS methodology separates our exercise dataset into five groups, and thus in this paper we will focus on this obtained solution for both approaches.

4.1. BOSS Model Approach

Table 1 depicts the distribution in age and gender of each cluster of exercises. As it can be observed, cluster 1 contains almost half of the exercises (41% of the total) and it is also the most heterogeneous in age and gender distribution. This cluster contains all the exercises performed by young and middle-aged male participants, but also a considerable amount of exercises coming from young women and almost one third of the exercises belonging to older adults. Cluster 5 contains also a important proportion of exercises, mostly performed by old female adults, but there is also a significant representation of younger females' exercises. The rest of the clusters are poorly represented by older adults and divide mainly exercises of young women into different groups. The information offered by this model does not provide enough knowledge about the characteristics of the exercises that are relevant to understand the obtained categorization.

Table 1. Cluster results for BOSS Model approach

Cluster	Age Y	Age M	Age O	Gender
c1	30	20	13	(42F/21M)
c2	14	8	0	(22F)
c3	9	3	3	(15F)
c4	4	4	1	(8F/1M)
c5	11	3	30	(38F/6M)

4.2. DTW - KL Divergence Approach

The DTW-KL Divergence approach returned a combination of five types of strides and five types of bags-of-strides. Figure 3 shows the shape of the medoids of each stride cluster in terms of $F_x diff$, where the bold line represents the medoid and the rest are its 10th closest strides, *i.e.* the most similar strides based on the shape of the pushing forces. The X-axis represents the stride time duration (in number of instances, each one every 100 ms). The Y-axis shows the force variation during the gait cycle in N. Figure 4 depicts the distribution of the bags-of-strides obtained in each Exercise cluster.

As it can be observed, strides are mainly grouped by its duration and shape. The first Exercise cluster is mainly represented by stride type 3, which contains the shorter strides in duration. These strides also present a balanced amount of force between both arms with values around 0 during the swing phase of the stride. Table 2 confirms that this group is mainly formed by young women (in this case 28 exercises performed by young female and two from a middle-age). This is coherent with the anthropometric characteristics of female humans: the stride length is strongly related to the individual's height, which is generally lower in women.

The second Exercise cluster characterises a mixed group, as in the BOSS model, with the higher number of observations. In this case, all the exercises executed by young and middle-aged men appear together. Surprisingly, the exercises of this cluster coming from female participants have the worst average performances of all age groups and clusters (*i.e.* women in this cluster performed at the lowest average gait velocity in comparison to women in same age category but different Exercise clusters). The older women in this group walk at low gait velocity, which is an indicator of risk, but according to the stride shape, they have a balanced interaction with the *i-Walker*, which probably means that they present cautious walking behaviour. This cluster is again the one containing more observations, being most of them of middle-aged people (26 out of 58). Since most of the strides are of type 2, it is likely to say it is the more general gait shape of all the strides collected in this dataset. People in this group are those presenting a minimal force variation. The gait shape indicates that a significant right-hand force was exerted during the phases regarding the right step, and then the forces applied from the two parts of the body get balanced to perform the left step. The stride shapes fall in the middle of stride length, being some of them very short (presumably, those performed by women). The force compensation in the swing phase of the gait cycle is quite balanced, with a general trend of applying more right forces, especially when reaching the end of the period.

The rest of Exercise clusters are mainly represented by the remaining stride types. The third Exercise cluster presents a combination of two types of strides (4 and 5 in Figure 3). Both strides present a considerable force variation when changing from right to left step in the gait cycle, specially in the last case. People executing these exercises

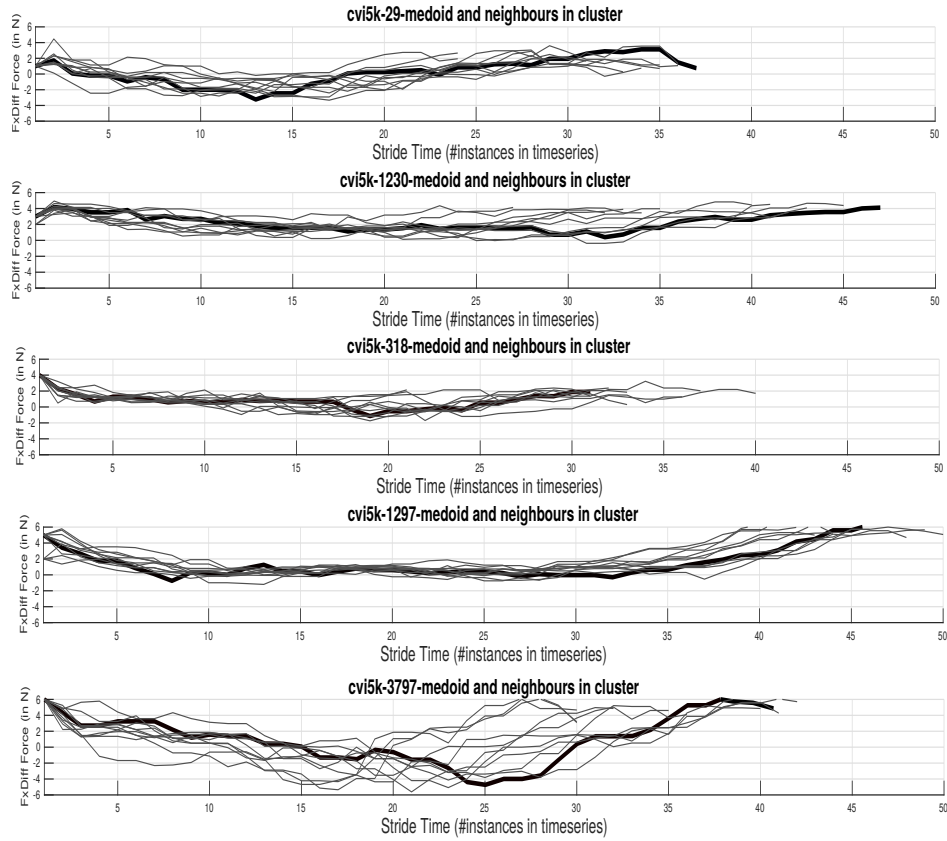


Figure 3. Gait shapes of each stride medoid cluster and its closest medoids/strides

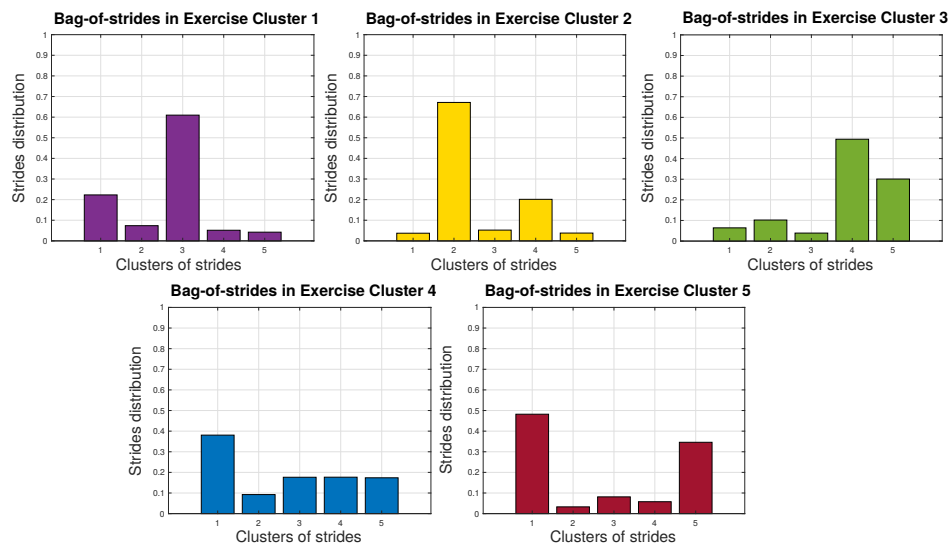


Figure 4. Cluster representation of bags-of-strides with five types of strides, grouped in five sorts of exercises.

Table 2. Anthropometric and spatio-temporal gait characteristics the CVI pilot

Cluster	1				2				3			
Age	N	Speed	Cadence	Stride Length	N	Speed	Cadence	Stride Length	N	Speed	Cadence	Stride Length
Age M	2	1,16	64,61	1,16	26	1,05	59,52	1,18	2	1,03	65,22	1,04
F	2	1,16	64,61	1,16	16	0,99	57,32	1,14	2	1,03	65,22	1,04
M					10	1,14	63,04	1,25				
Age O					13	0,50	54,13	0,58	11	0,88	61,33	0,93
F					13	0,50	54,13	0,58	8	0,85	60,00	0,91
M									3	0,97	64,88	0,97
Age Y	28	1,05	61,38	1,14	19	1,02	59,25	1,17	5	1,18	65,22	1,23
F	28	1,05	61,38	1,14	11	1,02	59,87	1,16	5	1,18	65,22	1,23
M					8	1,00	58,39	1,18				
Total	30	1,06	61,59	1,14	58	0,91	58,22	1,04	18	0,98	62,84	1,02
Cluster	4				5							
Age	N	Speed	Cadence	Stride Length	N	Speed	Cadence	Stride Length				
Age M	6	1,16	65,50	1,20	2	1,12	65,78	1,14				
F	6	1,16	65,50	1,20	2	1,12	65,78	1,14				
M												
Age O	12	0,84	57,46	0,97	11	0,68	50,26	0,75				
F	5	0,70	55,68	0,80	11	0,68	50,26	0,75				
M	7	0,95	58,74	1,09								
Age Y	7	1,23	64,84	1,31	9	1,02	60,80	1,10				
F	7	1,23	64,84	1,31	9	1,02	60,80	1,10				
M												
Total	25	1,03	61,46	1,12	22	0,86	55,98	0,93				

combine (a) long strides, with an increasing pushing force in the right step part of the cycle, but with a balanced phase during the swing phase with (b) shorter strides, were the individual uses a higher amount of force in both phases of the gait cycle. This should be analysed by an expert eye to determine if this is a possible identifier of gait disturbance or fall risk. The first type of stride, present in the fourth and fifth Exercise cluster is very similar in shape to the previously described, with the difference that the force variation during the swing phase is lower in this case. However, the trend to push with a significant amount of right force is also observable in this case. These three last clusters are mainly formed by older adults (clusters 3 and 4 contain the exercises performed by older males). However, we still find the presence of some exercises belonging to young or middle-aged women with high cadence and stride lengths, thus belonging to the high gait velocity.

5. Conclusions

This paper presents two approaches to gait analysis and human walking behaviour by clustering a group of exercises performed by individuals of different ages. The main difference relies on the treatment of the data before applying the machine learning techniques. Both methods use unsupervised learning techniques to analyse the dataset, with the challenge of determining the better solution.

The first approach uses a BOSS model to form a vocabulary of exercises, but provides only results in terms of the biological characteristics of the participants. However, the second approach uses a local maxima technique to identify the strides of the exercise, offering a better understanding on how do individuals interact with the *i-Walker*.

This method is able to extract the spatio-temporal characteristics of the exercise, offering means to learn how people drive the rollator in a narrow space and to detect certain anomalies, such as unbalanced amount of force on one side of the body, excessive use of the pushing force to perform the step or the variability of the strides within an exercise. All these characteristics could help experts to identify gait disturbances and, in a future, provide tailored navigation assistance through strategies of control.

This second approach is able to identify younger female participants, but also elders at higher risk of falling together. In addition, it complements the clustering result with anthropometric and spatio-temporal information of both the user and the exercise, which makes possible to provide insight knowledge about the interaction between older adults and the *i-Walker*, and thus to determine in a future which are their needs when walking and how the *i-Walker* can contribute to improve their quality of life. Further experimentation is needed with a gender-balanced trial, more focused on older adults presenting comorbidities or difficulties to walk without a rollator or any other assistive device.

References

- [1] L.Z. Rubenstein, Falls in older people: epidemiology, risk factors and strategies for prevention, *Age and Ageing* (2006) **35** s2 (ii34-ii41).
- [2] S. Fritz, M. Lusardi, White Paper: Walking Speed: the Sixth Vital Sign. *J Geriatric Physical Therapy* (2009) **32** 2 (2-5).
- [3] A.L. Rosso, et al., Aging, the Central Nervous System and Mobility. *J Gerontol A Biol Sci Med Sci* (2013)
- [4] M. Montero-Odasso, J. Verghese, O. Beauchet, J.M. Hausdorff, Gait and Cognition: A Complementary Approach to Understanding Brain Function and the Risk of Falling. *J American Geriatrics* (2012) **60** 11 (2127-2136).
- [5] P. Plummer-D'Amato, B. Brancato, M. Dantowitz, S. Birken, C. Bonke, E. Furey, Effects of Gait and Cognitive Task Difficulty on Cognitive-Motor Interference in Aging. *J Aging Research* (2012).
- [6] P. Schäfer, The BOSS is concerned with time series classification in the presence of noise. *Data Mining and Knowledge Discovery*, (2015), **29**, 1505-1530
- [7] World Health Organization, World report on ageing and health. *World Health Organization* (2015)
- [8] C. Prakash, R. Kumar, N. Mittal, Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artificial Intelligence Review* (2016)
- [9] M.W. Whittle, Gait analysis. *The Soft Tissues* (1993) 187-199.
- [10] A. Cortés, A. Martínez, J. Béjar, Using the *i-Walker* as a measurement tool for walking behaviour analysis. *4th Workshop on ICTS for Improving Patients Rehabilitation Research Techniques (REHAB)* (2016).
- [11] M. Ojeda, A. Cortés, J. Béjar, U. Cortés, Automatic classification of gait patterns using a smart rollator and the BOSS model. *11th Pervasive Technologies Related to Assistive Environments (PETRA)* (2018).
- [12] U. Cortés A.B. Martínez-Velasco, C. Barrué R. Annicchiarico, AI Based Fall Management Services - The Role of the *i-Walker* in I-DONTFALL. *Advances in Artificial Intelligence - 11th Mexican Int Conf on AI* (2012) Revised Selected Papers, Part I (395-406).
- [13] S. Sprager, M. Juric, Inertial Sensor-Based Gait Recognition: A Review, *Sensors* (2015) **15** 9.
- [14] M.M. Deza, E. Deza, Encyclopedia of Distances. *Springer* (2013).
- [15] J. Barth, et al., Stride Segmentation during Free Walk Movements Using Multi-Dimensional Subsequence Dynamic Time Warping on Inertial Sensor Data. *Sensors* (2015) **15** 3.
- [16] M.O. Derawi, P. Bours, K. Holien, Improved Cycle Detection for Accelerometer Based Gait Authentication. *2010 6th Int Conf Intelligent Information Hiding and Multimedia Signal Processing* (2010).
- [17] A. Pla, N. Mordvanyuk, B. López, M. Raaben, T.J. Blokhuis, H.R. Holstlag, Bag-of-steps: Predicting lower-limb fracture rehabilitation length by weight loading analysis. *Neurocomputing* (2017) **268**.
- [18] M. E. Tinetti, T. F. Williams, R. Mayewski. Fall risk index for elderly patients based on number of chronic disabilities. *The American journal of medicine* (1986) **80**(3), 429-434.