Discriminating Postural Control Behaviors from Posturography with Statistical Tests and Machine Learning Models: Does Time Series Length Matter?

Luiz H. F. Giovanini[1,2], Elisangela F. Manffra[1,3], and Julio C. Nievola[1,2]

[1] Pontifícia Universidade Católica do Paraná, Curitiba, Paraná, Brazil
[2] Programa de Pós-Graduação em Informática
[3] Programa de Pós-Graduação em Tecnologia em Saúde
{l.giovanini,elisangela.manffra,julio.nievola}@pucpr.br

Abstract. This study examines the influence of time series duration on the discriminative power of center-of-pressure (COP) features in distinguishing different population groups via statistical tests and machine learning (ML) models. We used two COP datasets, each containing two groups. One was collected from older adults with low or high risk of falling (dataset I), and the other from healthy and post-stroke adults (dataset II). Each time series was mapped into a vector of 34 features twice: firstly, using the original duration of 60 s, and then using only the first 30 s. We then compared each feature across groups through traditional statistical tests. Next, we trained six popular ML models to distinguish between the groups using features from the original signals and then from the shorter signals. The performance of each ML model was then compared across groups for the 30 s and 60 s time series. The mean percentage of features able to discriminate the groups via statistical tests was 26.5% smaller for 60 s signals in dataset I, but 13.5% greater in dataset II. In terms of ML, better performances were achieved for signals of 60 s in both datasets, mainly for similarity-based algorithms. Hence, we recommend the use of COP time series recorded over at least 60 s. The contribution of this paper also include insights into the robustness of popular ML models to the sampling duration of COP time series.

Keywords: Machine Learning, Artificial Intelligence, Feature Extraction, Posture, Posturography, Sampling Duration.

1 Introduction

Postural control (PC) is essential for the accomplishment of a variety of motor tasks and daily living activities [1]. The decline in this control - usually followed by aging or neurological diseases such as stroke - affects the mobility and independence, thus preventing the person from having a good quality of life. A practical way to characterize PC is through posturography, a technique that uses a device called force plate to record the body sway during quiet standing for a certain amount of time [1]. This sway is recorded as time series data of the center-of-pressure (COP) displacements of the person over its base of support in both x and y directions [1]. Then, with the help

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

of suitable metrics, COP time series can be parameterized into posturographic features able to work as clinical descriptors for many recognition tasks. Importantly, many widely-used metrics are influenced by the length of the COP time series [2], [3], which depends upon the sampling duration used for data recording. This is a critical point due to the lack of standardization of this acquisition parameter in posturography [4]. Some researchers claim that long durations of at least 120 s are necessary to fully characterize PC [3]. Conversely, some others criticize long durations arguing that factors such as fatigue can confound the results [5]. Hence, short durations are largely observed in the literature, usually around 30 s [2], [4], [5].

Traditionally, discrimination of COP behavior has been performed with statistical tests, where each posturographic feature is analyzed separately. More recently, some studies have successfully replaced such tests by ML models, where the discrimination is achieved by combining multiple features in a more sophisticated fashion. Two ways of COP discrimination are observed in the literature. The first one consists in comparing features from the same population group obtained at different balance tasks, thus helping understand the complexity of such tasks. This is known as intra-group analysis. The second way is the inter-group analysis, where researchers compare features derived from different groups aimed at discriminating them. This allows, for instance, assessing how different pathologies affect the PC.

Many posturographic metrics are influenced by the COP sampling duration, which typically ranges from 30 s to 60 s [4]. To the best of our knowledge, studies have dedicated to examine the sensitivity of such metrics to a variety of short durations for intra-group analyzes [2], [4]; however, similar investigations were not conducted yet for inter-group comparisons. As a first step in this direction, this paper aims at investigating the inter-group discriminative power of features computed from COP data of 30 s and 60 s for the use of both statistical tests and ML models. Since more accurate intra-group features have been reported for 60 s than 30 s [5], [6], we hypothesized that COP data of 60 s can also provide more discriminative inter-group features.

2 Methods

2.1 Datasets

We used two COP datasets, both recorded at quite standing over 60 s at a sample frequency of 100 Hz and filtered at 10 Hz (dual-pass 4th order low-pass Butterworth). Derived from a public database of older adults [7], dataset I has 864 instances (i.e., pairs of COPx and COPy time series), 432 from subjects with high risk of falling (ROF) and 432 from individuals with low ROF. We allocated a time series in the high ROF group when the individual fulfilled at least one of three main risk factors for falls in the elderly [8]: (i) history of falls in the past year; (ii) prevalence of fear of falling; (iii) a score smaller than 16 points at Mini Balance Evaluation Systems Test, which indicates significant balance impairments. Originally collected by [9], dataset II has 114 instances, 57 from post-stroke adults and 57 from healthy individuals. We have permission (no. 991.103) of the Ethics Committee of PUCPR to use such dataset

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

2.2 Feature Extraction

We implemented a Matlab routine to parameterize pairs of COP*x* and COP*y* time series into vectors of 34 features, which are displayed in Table 1. As shown, we included 13 magnitude metrics that derive from the overall size of the COP fluctuations, as well as 6 structural metrics to capture the temporal patterns in the COP dynamics [1], [5]. Out of these 19 metrics, 11 are temporal, 04 are spatial, and 04 are spectral. While temporal and spectral metrics are computed individually from the *x* and *y* directions of COP data, spatial metrics derive from both directions simultaneously [1], [5]. As can be seen, there are metrics derived from both displacement (COPd) and velocity (COPv) time series. For more information, including equations and implementation details, please refer to [1], [5]. To investigate our hypothesis, the feature extraction was performed twice for each dataset: firstly, using the original time series of 60 s (6000 data points), and then truncating them in the first 30 s (the first 3000 points).

Table 1. Summary of metrics used for COP parameterization.

Category	Type	Metrics
Magnitude	Temporal	Mean distance, root mean square (RMS) distance, mean velocity, RMS velocity, standard deviation (SD) of velocity.
	Spatial	Sway path, length of COP path, excursion area, total mean velocity.
	Spectral	Mean frequency, median frequency, Fp% of spectral power (p=80, 95).
Structural	Temporal	Sample entropy (SE) of distance, SE of velocity, multiscale sample entropy (MSE) of distance, MSE of velocity, scaling exponent of velocity, Hurst exponent of distance.

All magnitude features were computed after removing the offset of the COPd signals by subtracting the mean [1]. The spectral features were calculated via Welch's periodogram method with a Hamming window with 50% of overlap [5]. Prior to the SE and MSE analyses, in order to remove nonstationarities and long-range correlations that may confound results, we detrended the COPd signals via Empirical Mode Decomposition method by subtracting from signals the four last Intrinsic Mode Functions of lowest frequency (0.05 Hz to 1Hz) [10]. Then, we calculated SE taking $N = 2$ and $r = 0.15$ for COPd [10] and $N = 2$ and $r = 0.55$ for COPv [5], where N is the number of data points and r is the tolerance threshold. The scaling exponent (α) and Hurst exponent (H) were computed, respectively, via Detrend Fluctuation Analysis (DFA) and Scaled Windowed Variance (SWV) methods. We computed α from COPv signals only, and H from COPd signals only [11].

2.3 Machine Learning Experiments

For pattern recognition, we considered six popular ML models with specific configurations successfully used by past works to handle COP features [11], [12]: *k*-Nearest Neighbors (*k*-NN) with $k = 1, 3, 5, \ldots, 19$; Decision Tree unpruned (DT1) and pruned (DT2); Multilayer Perceptron with 500-epochs training time and 0% validation set

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

size (MLP1), 10 thousand-epochs and 5% validation size (MLP2), and 10 thousand-epochs and 10% validation size (MLP3); Naïve Bayes (NB); Random Forest (RF) with six features used in random selection; Support Vector Machines with 3rd degree RBF kernel and cost 1 (SVM1) and cost 10.0 (SVM2). For each dataset, the input features were normalized to a 0-1 range. Then, using the Weka software, the learning algorithms were trained and tested within 10 repetitions via 10-fold cross-validation for dataset I, and via leave-one-out for dataset II due to the small number of instances. As both datasets are balanced, we adopted the accuracy as performance metric. Each algorithm was trained and tested under each dataset twice: firstly, using the features computed from original COP time series of 60 s, and then using the features calculated from shorter signals of 30 s.

2.4 Statistical Analyses

Firstly, for each dataset, we performed an intra-group analysis where each feature was compared across original (60 s) and shortened (30 s) COP time series using the Wilcoxon test. Next, using the Mann-Whitney U-Test, we conducted an inter-group analysis of each feature for both original and shortened data. Lastly, to analyze the influence of the sampling duration on the ML models, the accuracy of each learning algorithm was compared across 60 s and 30 s features via Mann-Whitney U-Test. Using the same test, we also compared the global mean accuracies computed over all models. The level of confidence adopted was 95%. The normality of all results was verified via Lilliefors test. These analyzes were conducted by using the Matlab R2013b.

3 Results and Discussion

3.1 Intra- and Inter-Group Sampling Duration Effects

Table 2 displays the statistical results of our intra-group analysis, where most features have shown to be sensitive to the decreasing of the sampling duration. Similar results were reported by past studies dedicated to address the question of optimal sampling duration for COP data acquisition. For example, after examining COP data recorded over 15, 30, 60, and 120 s from healthy young adults, [6] concluded that longer durations of at least 60 s are necessary to ensure more reliable RMS distance and mean frequency features in an intra-group analysis. A similar conclusion was drafted by [4], [5] based on a variety of magnitude and structural COP features. All these findings corroborate that, when performing either intra- or inter-group analyzes from COP data, comparisons should be limited to features calculated from samples of equal duration, otherwise they may lead to misinterpretations [6].

Table 2 also shows the statistical results of our inter-group analysis. To the best of our knowledge, this is the first study to report the sampling duration effects on the discriminative power of COP features on older adults with low or high ROF as well as on healthy and post-stroke adults. Surprisingly, our results provided contrasting conclusions for these population groups. While the mean percentage of discriminative features grown 26.5% with the decreasing of the sample duration for dataset I, it de-

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

creased 13.5% for dataset II. In other words, the ROF was considerably better recognized from COP time series of 30 s, whereas the contrasts in PC between healthy and post-stroke volunteers were more detectable when using 60 s COP signals. In summary, as these findings allow us accepting our hypothesis for dataset II only, we concluded that the optimal sampling duration in terms of discriminative features depends upon the populations under analysis. Hence, it seems advisable to record COP data over at least 60 s, as argued by other studies [5], [6], and then truncate the signals to examine the optimal sampling duration in each case.

Table 2. Statistical values obtained in both intra- and inter-group analyzes.

Feature	Intra-group p-values				Inter-group p-values			
	Dataset I		Dataset II		Dataset I		Dataset II	
	High ROF	Low ROF	Stroke	Healthy	60 s	30 s	60 s	30 s
Mean distance	★	★	n.s.	**	n.s.	n.s.	n.s.	n.s.
RMS distance	★	★	n.s.	**	n.s.	*	n.s.	n.s.
Mean velocity	*	**	★	★	n.s.	n.s.	★	★
RMS velocity	*	**	★	★	n.s.	n.s.	★	★
SD of velocity	n.s.	**	**	★	n.s.	n.s.	★	★
Sway path	★	★	★	★	n.s.	n.s.	★	★
Length of COP path	★	★	★	★	n.s.	n.s.	n.s.	n.s.
Excursion area	★	★	n.s.	★	n.s.	n.s.	n.s.	n.s.
Total mean velocity	★	★	★	★	n.s.	n.s.	★	★
Mean frequency	★	★	★	★	*	**	★	**
Median Frequency	★	★	★	★	n.s.	**	**	**
F80%	★	★	*	★	*	★	★	**
F95%	★	★	★	★	*	★	★	**
SE of distance	★	★	★	★	n.s.	★	**	*
SE of velocity	★	★	★	★	*	*	n.s.	n.s.
MSE of distance	★	★	★	★	n.s.	★	★	**
MSE of velocity	★	★	★	★	*	★	n.s.	n.s.
Scaling exponent	★	★	★	★	*	★	**	**
Hurst exponent	★	★	★	★	*	★	**	n.s.
Percentage of $p < 0.05$	90.0	92.5	78.1	95.0	17.7	44.2	59.8	46.3

The *, **, and ★ symbols denote, respectively, $p < 0.05$ for COP data in x direction only, y direction only, and both directions. n.s. means not significant ($p \geq 0.05$) for both directions.

3.2 Sampling Duration Effects on the Machine Learning Results

Table 3 shows the influence of the COP sampling duration on the accuracy of the ML models trained in this work. From a general perspective, the original COP time series

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

yielded slightly better global accuracies than the shortened signals. These results suggest that a sampling duration of 60 s provides more discriminative information than 30 s when distinguishing groups via popular ML models, thus supporting our hypothesis. One should notice, however, that the global accuracies were manly influenced by the performance of *k*-NN, especially in the case of dataset I. Conversely, some learning algorithms have shown robustness to the COP duration: DT2, MLP2, MLP3, NB, and SVM2. Based on these findings, it is possible to infer that similarity-based ML methods such as *k*-NN are more sensitive to the sampling duration than other popular models. Thus, they should be avoided in certain situations, for example, when dealing with COP time series recorded over too short durations that prevent good results, or when trying to distinguish populations whose COP data were recorded over different durations. Otherwise, one must be careful to identify how much performance is driven by the PC behaviors under analysis and how much is a function of COP duration.

Table 3. Machine learning results.

Model	Mean accuracy (%) for dataset I		Mean accuracy (%) for dataset II	
	60 s	30 s	60 s	30 s
1-NN	61.8	61.2	**66.7**	60.5
3-NN	**64.1**	60.3	66.7	66.7
5-NN	**62.8**	60.0	68.4	68.4
7-NN	**63.3**	59.8	72.8	69.3
9-NN	**62.6**	59.9	**71.1**	64.9
11-NN	**63.2**	58.8	**71.1**	65.8
13-NN	**62.8**	59.0	72.8	71.1
15-NN	**62.8**	59.8	71.1	68.4
17-NN	**63.0**	60.1	69.3	69.3
19-NN	**62.3**	60.2	66.7	69.3
DT1	57.0	56.0	57.9	**64.9**
DT2	57.1	56.0	61.4	64.9
MLP1	**61.7**	58.7	65.4	62.8
MLP2	58.7	57.3	63.9	61.6
MLP3	59.0	57.3	64.8	61.8
NB	58.4	58.3	68.4	67.5
RF	**64.9**	61.0	71.9	70.6
SVM1	58.2	57.4	**67.5**	63.2
SVM2	60.1	60.0	71.1	67.5
Global mean	**61.3**	59.0	**67.8**	66.2

Statistically ($p < 0.05$) greater accuracies are marked in bold.

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28

4 Conclusion, Future Work, and Acknowledgment

This paper examined the effects of COP short durations of 30 s and 60 s on the discriminative power of posturographic features in inter-group comparisons using statistical tests and popular ML models. Conclusions are limited to the population groups analyzed here: older adults with high or low ROF, healthy and post-stroke adults. In terms of statistical tests, we concluded that the optimal COP duration changes according to the group under analysis. However, when using ML, COP signals of 60 s have proved to be more discriminative, mainly for similarity-based models. Therefore, we advise one recording COP data over at least 60 s, and then truncating the time series if necessary, depending on the tools to be employed or questions to be investigated. To ensure the repeatability of the experiments performed in this work, we made available to download our COP features and Matlab codes at https://goo.gl/TACWYt. Future work will focus on improving ML performance by testing models of the state-of-the-art for time series classification, such as convolutional and recurrent neural networks.

L. H. F. Giovanini is thankful to PUCPR for his scholarship. We would like to thank NVIDIA Corporation for the donation of a Titan X Pascal GPU.

References

1. Duarte, M. & Freitas, S. M. Revision of posturography based on force plate for balance evaluation. Brazilian Journal of physical therapy 14, 183–192 (2010).
2. Rhea, C. K., Kiefer, A. W., Wright, W. G., Raisbeck, L. D. & Haran, F. J. Interpretation of postural control may change due to data processing techniques. Gait & posture 41, 731–735 (2015).
3. van der Kooij, H., Campbell, A. D. & Carpenter, M. G. Sampling duration effects on centre of pressure descriptive measures. Gait & posture 34, 19–24 (2011).
4. Ruhe, A., Fejer, R. & Walker, B. The test–retest reliability of centre of pressure measures in bipedal static task conditions–a systematic review of the literature. Gait & posture 32, 436–445 (2010).
5. Kirchner, M., Schubert, P., Schmidtbleicher, D. & Haas, C. T. Evaluation of the temporal structure of postural sway fluctuations based on a comprehensive set of analysis tools. Physica A: Statistical Mechanics and its Applications 391, 4692–4703 (2012).
6. Carpenter, M. G., Frank, J. S., Winter, D. A. & Peysar, G. W. Sampling duration effects on centre of pressure summary measures. Gait & posture 13, 35–40 (2001).
7. Santos, D. A. & Duarte, M. A public data set of human balance evaluations. PeerJ Preprints (2016). doi:https://doi.org/10.7287/peerj.preprints.2162v1
8. Organization, W. H. WHO global report on falls prevention in older age. (World Health Organization, 2008).
9. Silva, S. M. Análise do controle postural de indivíduos pós-acidente vascular encefálico frente a perturbações dos sistemas visual e somatossensorial. (PUCPR, 2012).
10. Costa, M. et al. Noise and poise: enhancement of postural complexity in the elderly with a stochastic-resonance–based therapy. EPL (Europhysics Letters) 77, 68008 (2007).
11. Giovanini, L. H., Silva, S. M., Manffra, E. F. & Nievola, J. C. Sampling and Digital Filtering Effects When Recognizing Postural Control with Statistical Tools and the Decision Tree Classifier. Procedia Computer Science 108, 129–138 (2017).
12. Goh, K. L. et al. Typically developed adults and adults with autism spectrum disorder classification using centre of pressure measurements. in 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 844–848 (IEEE, 2016).

ICCS Camera Ready Version 2018
To cite this paper please use the final published version:
DOI: 10.1007/978-3-319-93713-7_28