Feature Extraction using Poincaré Plots for Gait Classification

Luís Marques[1]
lf1119@ic.ac.uk
Flora Ferreira[2]
flora.ferreira@gmail.com
Aldina Correia[3]
aic@estg.ipp.pt
Estela Bicho[4]
estela.bicho@dei.uminho.pt
Wolfram Erlhagen[2]
wolfram.erlhagen@math.uminho.pt

[1] Department of Aeronautics, Imperial College London
[2] Center of Mathematics, University of Minho
[3] Center for Research and Innovation in Business Sciences and Information Systems, ESTG, Polytechinc Institute of Porto
[4] Algoritmi Center, University of Minho

Abstract

The aim of this study is to evaluate different features, extracted from a Poincaré plot of gait signals, in their ability to classify the gait of patients with neurodegenerative diseases: Parkinson's disease (PD) and Huntington's disease (HD). Five different features that describe gait variability were extracted from the Poincaré plots of two gait signals: stride time and percentage of stride time spent in swing phase. Among the set of extracted features, those that displayed significant differences between the two groups and were not correlated with each other, were used as input to the support vector machine classifier. It was found that all extracted features (with exception of one feature in PD vs healthy group comparison) are significantly different between healthy and pathological subjects and are suitable to discriminate them (with accuracies greater than 80%). When comparing PD vs HD, just three features were significantly different, however, a relatively good classification accuracy (around 72%) was achieved using two of them. The results demonstrate that it is feasible to apply variability measures extracted from Poincaré plots of gait data signals in gait classification problems.

1 Introduction

There has been a growing trend towards the analysis of the human gait signals as a complementary tool to the diagnosis and management of patients with neurodegenerative diseases [1, 3, 7]. Gait stride-to-stride variability is an important marker to better understand the mechanisms of movement disorders. Increased stride variability has been associated with an increase of fall risk in older adults in general [6]. One of the major objectives of studying gait patterns is to identify features, from gait time series, that show signs of gait disturbances due to aging or neurodegenerative diseases, such as Parkinson's disease (PD) and Huntington's disease (HD). Poincaré plots have been widely used in the cardiovascular area to measure heart rate variability and their application on gait time series is emerging [2, 4]. Poincaré plot is a geometrical visual representation of the time series that describes the evolution of the series and that allows for the assessment of series time dynamics. Khandokar et al. [4] quantified gait dynamics in elderly subjects using Poincaré plot measures of minimum foot clearance variability and found that, by monitoring these measures, it is possible to improve the gait performance of falls risk. Golińska [2] used Poincaré plot's standard descriptors, namely SD_1, SD_2 and SD_1/SD_2 ratio, to distinguish healthy subjects from patients with PD and HD, and observed that too low or too high SD_1/SD_2 ratio values could be connected with disease, however, further research is needed. To go beyond the standard descriptors, we have used other measures based on PP (such as SD_{UP} and SD_D described below) that represent different aspects of gait dynamics. In this study, the aim is to evaluate different gait variability measures derived from the Poincaré plot in their ability to classify the gait of patients with neurodegenerative diseases.

2 Materials and methods

2.1 Gait data

The gait database in this study was provided by Hausdorff et al. [3] and it is available at https://physionet.org/physiobank/database/gaitndd/. The gait signals were recorded when subjects walked at their normal pace along a straight hallway of 77 meters. Stride-to-stride measures of footfall contact times were derived from these original signals. The original database consists of 15 PD subjects (5 female, age range [44,77] years), 20 HD subjects (14 female, age range [29,71] years), and 16 healthy CO subjects (14 female, age range [22,74] years). However, one female HD patient, who was 33 years old, was not included in this study because the values of the stride intervals for her right foot are constantly a fixed value higher than the elapsed time between the two consecutive strides. Heights and weights were not significantly different between the three groups.

2.2 Poincaré Plot and feature extraction

Poincaré plot is a scatterplot of the current gait cycle against the preceding gait cycle [2] (panels C) and D) of Fig.1). For a given time series $x_0, x_1, x_2, ..., x_n$ the Poincaré plot is a plot of points (x_0, x_1), (x_1, x_2),..., (x_{n-1}, x_n). Variability of data, along the width and length of the short and long axis of the ellipse, that better fits the data, is represented by short term (SD_1) and long term (SD_2) variability. Measures SD_1 and SD_2 can be defined as:

$$SD_1 = \frac{\sqrt{2}}{2} SD(x_n - x_{n+1}) \tag{1}$$

and

$$SD_2 = \sqrt{2SD(x_n)^2 - \frac{1}{2} SD(x_n - x_{n+1})^2} \tag{2}$$

where SD is a standard deviation. Another measure is obtained with the ratio between SD_1 and SD_2, SD_1/SD_2. Comparing the plots in Fig. 1, one can find that the dispersion of points for the PD patient exhibits greater SD_1 and SD_2 values. Additionally, the lowest of SD_1/SD_2 ratio value appears on the healthy subject. The variability above and below the identity (I_d) line are represented by SD_{UP} and SD_D, respectively [5]. These two variability measures are obtained by calculating the distance of each point to the I_d line, D_{UP} for points above and D_D for those below, according to the equations (3) and (4)

$$SD_{UP} = \frac{1}{n_{UP}} \sum_{i=1}^{n_{UP}} (D_{UP_i})^2 \tag{3}$$

$$SD_D = \frac{1}{n_D} \sum_{i=1}^{n_D} (D_{D_i})^2 \tag{4}$$

where n_{UP} and n_D are the number of points in each case.

In this study, two gait characteristics were used: stride time and percentage of stride time spent in swing phase (swing (% of stride)), both using as reference the right foot. Stride time is the time between consecutive heel strikes of the same foot. Each stride has two phases: stance and swing. The swing phase is the part of the stride during which the reference foot does not contact the ground, which is normally about 40% of stride time. For each gait characteristic the following features were calculated: SD_1, SD_2, SD_1/SD_2, SD_{UP} and SD_D, having obtained 10 measures/features for each subject.

2.3 Statistical Analysis and Classification

Statistical analysis was used to identify measures that are significantly different between two groups: PD vs. CO, HD vs. CO, and PD vs. HD. Mann–Whitney U test was used to compare the extracted features

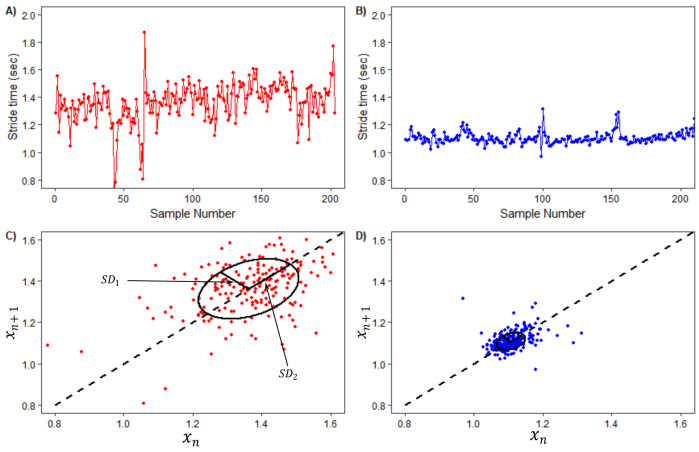

Figure 1: Right-foot gait stride-interval time series of a 64-year-old Parkinson's Disease patient and a 61-year-old healthy subject and its corresponding Poincaré plots.

for each two groups. The features with statistical significance (p-values < 0.05) were selected. Then, Pearson correlation coefficient (r) was used to measure the strength of a linear association between the features, and when high correlation ($r \geq 0.7$) was found between two features, the feature with higher p-value was discarded. For each binary classification, a linear kernel support vector machine (SVM) was used. SVM has been shown to be a powerful tool for solving classification problems based on gait with superior classification performance [7]. A 10-fold cross-validation was carried out to evaluate the performance of the SVM classifiers. Each dataset was randomly divided into 10 subsets, such that each subject's data was all contained within the same subset. The averages of the metrics (Accuracy, Specificity, and Sensitivity) are used to evaluate the classification's performance. Statistical analysis and classifications were performed using R statistical and computing software version 3.3.2 (http://www.rproject.org, R Foundation for Statistical Computing, Vienna, Austria).

3 Results

From Mann-Whitney analysis, the number of features that presented statistically significance was: 9 (expect SD_1/SD_2 of swing (% of stride)) when comparing PD patients and CO group, 10 (all features) when comparing HD patients and CO group, and just 3 (SD_1/SD_2, SD_D and SD_1) when comparing PD and HD. In each case, after applying the correlation analysis and discarding the correlated features, a reduced number of features was selected for the next stage. Selected features are reported in Table 1.

PD vs CO	HD vs CO	PD vs HD
SD_2 of SW	SD_{UP} of ST	SD_1/SD_2 of SW
SD_{UP} of ST	SD_1 of SW	SD_D of SW
SD_1/SD_2 of ST	SD_1/SD_2 of ST	
	SD_1/SD_2 of SW	

Table 1: Selected features (ST:Stride time, SW:Swing (% of stride)).

The means and standard deviations for selected features in three groups are reported in Table 2. The lowest gait variability measures are observed for the healthy subjects while the highest measures were observed for the HD group.

The performance of SVM classifications, using as input the selected features presented in Table 1, is reported in Table 3. The best results were obtained when comparing a group of patients, PD or HD, with the CO group, with accuracies above 83%.

4 Conclusions and Future Work

In recent years, different features based on the Poincaré plot have been defined and each of them can represent different aspects of gait dynamics. In this study, conventional (SD_1, SD_2 and SD_1/SD_2) and new features (SD_{UP} and SD_D, [5]) extracted from Poincaré plot of gait time series were analyzed in order to identify features which were significantly different between groups. Furthermore, selected features were used as an input for the SVM classifier for three different binary gait classifications across three groups: PD, HD, and CO. The results show that by using a reduced number of features extracted from a Poincaré plot of gait time series, a relatively high prediction performance can be achieved with an SVM classifier, when it comes to the classification between a group of patients, PD or HD, with healthy subjects, and also between between PD and HD patients. This study also confirms that both PD and HD display higher gait variability when compared to healthy subjects, while PD and HD may share more common features.

In conclusion, the features extracted from Poincaré plots of gait data signals could be useful in the classification of patients with neurodegenerative diseases. The promising results of this research warrant future study with the inclusion of other features extracted from Poincaré plot (e.g. Complex Correlation Measure), with additional gait time series (e.g., double support duration) and with different and larger data sets.

Feature	CO	PD	HD
SD_{UP} of ST	0.001 ± 0.001	0.181 ± 0.650	0.038 ± 0.064
SD_1/SD_2 of ST	0.588 ± 0.122	0.768 ± 0.237	0.806 ± 0.162
SD_1 of SW	1.266 ± 0.447	2.494 ± 0.617	4.305 ± 2.676
SD_2 of SW	1.590 ± 0.581	3.232 ± 0.897	4.367 ± 2.235
SD_1/SD_2 of SW	0.805 ± 0.115	0.783 ± 0.115	0.947 ± 0.174
SD_D of SW	1.828 ± 1.307	6.466 ± 3.116	26.07 ± 31.85

Table 2: Mean and standard deviation of selected features across groups.

PD vs. CO			HD vs. CO			PD vs. HD		
Acc	Sn	Sp	Acc	Sn	Sp	Acc	Sn	Sp
84%	90%	80%	83%	75%	95%	74%	80%	70%

Table 3: Performance measures, Accuracy (Acc), Sensitivity (Sn) and Specificity (Sp), obtained in each binary classification.

References

[1] C. Fernandes, L. Fonseca, F. Ferreira, M. Gago, L. Costa, N. Sousa, C. Ferreira, J. Gama, W. Erlhagen, and E. Bicho. Artificial neural networks classification of patients with parkinsonism based on gait. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2024–2030. IEEE, 2018.

[2] A. K. Golińska. Poincaré plots in analysis of selected biomedical signals. *Studies in logic, grammar and rhetoric*, 35(1):117–127, 2013.

[3] J. M. Hausdorff, A. Lertratanakul, M. E Cudkowicz, A. L. Peterson, D. Kaliton, and A. L. Goldberger. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. *Journal of applied physiology*, 88(6):2045–2053, 2000.

[4] A. H. Khandoker, M. Palaniswami, and R. K. Begg. A comparative study on approximate entropy measure and poincaré plot indexes of minimum foot clearance variability in the elderly during walking. *Journal of neuroengineering and rehabilitation*, 5(1):4, 2008.

[5] J. Rodriguez, A. Voss, P. Caminal, A. Bayés-Genis, and B. F. Giraldo. Characterization and classification of patients with different levels of cardiac death risk by using poincaré plot analysis. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1332–1335. IEEE, 2017.

[6] J. Verghese, R. Holtzer, R. B. Lipton, and C. Wang. Quantitative gait markers and incident fall risk in older adults. *The Journals of Gerontology: Series A*, 64(8):896–901, 2009.

[7] H. Zheng, M. Yang, H. Wang, and S. McClean. Machine learning and statistical approaches to support the discrimination of neurodegenerative diseases based on gait analysis. In *Intelligent patient management*, pages 57–70. Springer, 2009.