Classification of Early-Mild Subjects with Parkinson's Disease by Using Sensor-Based Measures of Posture, Gait, and Transitions

Luca Palmerini[1], Sabato Mellone[1], Guido Avanzolini[1], Franco Valzania[2], and Lorenzo Chiari[1]

[1] Department of Electrical, Electronic and Information Engineering, DEI, University of Bologna, Viale Risorgimento 2, 40136 Bologna, Italy
{luca.palmerini,sabato.mellone,guido.avanzolini, lorenzo.chiari}@unibo.it
[2] Department of Neuroscience, University of Modena and Reggio Emilia, via Pietro Giardini 1355, 41126 Baggiovara (MO)
f.valzania@ausl.mo.it

Abstract. Evaluation of posture, gait, turning, and different kind of transitions, are key components of the clinical evaluation of Parkinson's disease (PD). The aim of this study is to assess the feasibility of using accelerometers to classify early PD subjects (two evaluations over a 1-year follow-up) with respect to age-matched control subjects. Classifying PD subjects in an early stage would permit to obtain a tool able to follow the progression of the disease from the early phases till the last ones and to evaluate the efficacy of different treatments. Two functional tests were instrumented by a single accelerometer (quiet standing, Timed Up and Go test); such tests carry quantitative information about impairments in posture, gait, and transitions (i.e. Sit-to-Walk, and Walk-to-Sit, Turning). Satisfactory accuracies are obtained in the classification of PD subjects by using an *ad hoc* wrapper feature selection technique.

Keywords: Classification, Feature Selection, Parkinson's disease, Accelerometer.

1 Introduction

Evaluation of posture, gait, turning, and different kind of transitions, are key components of the clinical evaluation of Parkinson's disease (PD). The aim of this study is to assess the feasibility of using inertial sensors (accelerometers) to discriminate (classify) early PD subjects with respect to age-matched control subjects. Classifying PD subjects in an early stage would permit to obtain quantitative information that could be used as a decision-support system for the clinician, in a perspective of evidence-based medicine. This is the preliminary step in order to obtain a tool able to follow the progression of the disease from the early phases to the last ones and to evaluate the efficacy of different treatments. This is why in this study also

N. Peek, R. Marín Morales, and M. Peleg (Eds.): AIME 2013, LNAI 7885, pp. 176–180, 2013.

a 1-year follow-up was considered: the aim was to find a method of classification that would be accurate and robust over time.

Two tests were instrumented by a single accelerometer (quiet standing - QS, instrumented Timed Up and Go test - iTUG).

Quiet standing can provide quantitative information about postural sway and postural tremor [1]. The iTUG [2-4] is the instrumented version of a simple clinical test already used by clinicians to evaluate the locomotor performance of the elderly and PD subjects; this instrumented test can provide quantitative information on possible impairments in gait and transitions (i.e., Turning, Sit-to-Walk, Walk-to-Sit).

The two tests were already considered in previous works ([1], [2]) to quantify postural [1] and locomotor [2] impairments separately with no evaluation in follow-ups.

2 Methods

We examined 20 early-mild PD subjects) and 20 healthy age-matched control subjects (CTRL, 64±6 years old, 7 males and 14 females). Thirteen PD subjects also did a 1-year follow-up. All PD subjects were examined OFF medication (Hoehn & Yahr stage ≤ 3, 62±7 years old, 12 males and 8 females. The OFF condition in PD subjects was obtained by a medication washout of at least 18 hours. Subjects wore a tri-axial accelerometer, McRoberts© Dynaport Micromod, on the lower back at L5 level. They performed QS trials [1] in 5 different conditions (eyes open/closed, dual task, standing on foam with eyes open/closed) and iTUG trials [2]. The dual task consisted in counting audibly backwards from 100 by 3s. The QS trial consisted in standing quietly for 30 seconds with arms crossed on the chest; the iTUG trial consisted of rising from a chair (Sit-to-Walk), walking 7m at preferred speed (Gait), turning around (Turning), returning and sitting down again (Walk-to-Sit).

Fig. 1. Quiet Standing and Instrumented Timed Up and Go

The acceleration signals were recorded for each subject along the three orthogonal axes of the accelerometer: the first aligned with the direction of gait progression and coincident with the biomechanical anteroposterior (AP) axis of the body (front-back);

the second in the left/right direction and coincident with the biomechanical mediolateral (ML) axis of the body, and the third in the vertical direction.

Several measures were extracted from the acceleration signals.

The acceleration-derived measures extracted from QS trials quantify tremor and postural sway from quiet standing (e.g. power of high frequencies, root mean square of the signal, sway area…) [1].

The acceleration-derived measures from iTUG trials quantify duration of different components of the test, smoothness and variability of gait, range of motion during transitions (i.e. Sit-to-Walk, Walk-to-Sit), and so on [2]. The components analyzed for this study were: Sit-to-Walk, Gait, and Walk-to-Sit; Turning will be considered in following studies.

Only measures which could be considered reliable were kept for the feature selection procedure (as explained in [2]).

Finally, the total number of measures (features) which were considered is 27 (18 from QS, 9 from iTUG).

2.1 Feature Selection

Feature selection was applied in order to improve the performance of the classifiers. To select, from all the available features, a *wrapper* [5] feature selection was implemented, which was designed in [2]. To classify between the two groups we used the linear discriminant analysis (LDA) because we wanted a simple an easily interpretable classifier, to permit a clinical interpretation of the result.

In the feature selection procedure an exhaustive search among subsets of cardinality from one to three was performed. The limit of three was chosen to permit a clinical interpretation of the result (it would be difficult to associate too many features with different aspects of the disease).

Since feature selection is part of the tuning design of the classifier, it needs to be performed on the training set, in order to avoid overfitting in the final evaluation of the accuracy of the classifier [5].

The available data samples are 53: 20 CTRL subjects, 20 PD subjects on their first evaluation, and 13 PD subjects (a subset of the original 20) on their follow-up.

We randomly splitted the data in "70%-30% training-testing" by keeping the same proportion of CTRL and PD subjects in training and testing sets.

The 70%-30% rule was also applied to keep the 70% of PD subjects who had the follow-up evaluation in the training and the remaining 30% in the testing set.

Therefore, a PD subject with the follow-up had both the first evaluation sample and the follow-up sample either in the training or in the testing set. This was done in order to avoid overlapping of the datasets.

Finally, we considered 37 samples in the training set:

- 14 CTRL subjects (only first evaluation) = 14 CTRL samples
- 9 PD subjects * 2 (both first evaluation and f-up) = 18 PD samples
- 5 PD subjects (only first evaluation) = 5 PD samples

and 16 samples in the testing set:

- 6 CTRL subjects (only first evaluation) = 6 CTRL samples
- 4 PD subjects * 2 (both first evaluation and f-up) = 8 PD samples
- 2 PD subjects (only first evaluation) = 2 PD samples

The feature selection was performed in the training set with a leave-10%-out cross validation.

The classifier, built with the subset selected in the training set, was then tested in the testing set.

The LDA classifier was also tested on the same testing set with two benchmarks in order to evaluate whether the feature selection procedure that we implemented improves the performance of the selected classifier**.** The two benchmarks were:

- no feature selection;
- Principal Component Analysis: principal components that explained 90% of the data were selected.

3 Results and Discussion

In Table 1, the estimated misclassification rates obtained with the different FS methods are presented together with corresponding selected subsets. It can be seen that the best accuracy was achieved by the subset selected by the *wrapper* technique. Two objective features extracted from iTUG which quantify gait dynamics and smoothness, together with the total duration of the test, can discriminate with a satisfactory accuracy between early-mild PD and CTRL subjects. This was obtained with a simple classifier and few features: a possible clinical interpretation of the results is that lateral dynamics (range of motion) and vertical smoothness (reproducibility of step patterns) during gait are already impaired in early-mild PD. In addition, total duration of the iTUG test, even if it is not significantly different between CTRL and PD subjects [2] can help improving the classification. From the results it seems that iTUG is a better test with respect to QS (no features from QS were selected) in order to have an accurate classification which is also robust over

Table 1. Results of the classification with the LDA classifier

FS Method	Best Subset	Accuracy % [CI]	Sens %	Spec %
Wrapper	Total Duration of the iTUG Lateral gait dynamics during iTUG Vertical gait smoothness during iTUG	93.75 [72-99]	100	83.3
PCA	11 pcs	68.75 [44-86]	80	50
None	All Features	31.25 [84-86]	70	66.7

time. It is possible however that by considering more than three features, and/or changing classifier/FS method, accuracy could be improved by considering also QS features. The obtained accuracy would not have been obtained without feature selection; in fact considering all the features altogether would lead to a lower accuracy. This reflects the importance of performing feature selection in this kind of datasets.

It has to be noted that our relatively small sample size limits the power of our data mining perspective, as it can be seen by the large confidence intervals of the accuracies; however, separate training and testing sets were considered in order to obtain generalizable results.

4 Conclusion

The main result achieved by this work is a set of few quantitative measures, derived from a clinical test for locomotor evaluation, which can discriminate with a good accuracy between early-mild PD (both at their first evaluation and at 1-year follow-up) and CTRL motor patterns

Further experiments should be made on new subjects to validate these findings; it should also be investigated whether the presented measures remain valid for later stages of the disease and if they can track the evolution of the severity of symptoms.

Acknowledgments. The research leading to these results has received funding from the European Union - Seventh Framework Programme (FP7/2007-2013) under grant agreement n°288516 (CuPiD project). The authors wish to thank Luca Codeluppi, MD, and Valentina Fioravanti, MD, from the Department of Neuroscience, University of Modena and Reggio Emilia, Modena, Italy, for clinical supervision and assistance in data.

References

1. Palmerini, L., Rocchi, Mellone, S., Valzania, F., Chiari, L.: Feature selection for accelerometer-based posture analysis in Parkinson's disease. IEEE Transactions on Information Technology in Biomedicine 15(3), 481–490 (2011)
2. Palmerini, L., Avanzolini, G., Mellone, S., Valzania, F., Chiari, L.: Quantification of Motor Impairment in Parkinson's Disease Using an Instrumented Timed Up and Go Test. IEEE Transactions on Neural Systems and Rehabilitation Engineering (preprint, 2013)
3. Weiss, A., Herman, T., Plotnik, M., Brozgol, M., et al.: Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease? Med. Eng. & Phys. 32(2), 119–125 (2010)
4. Zampieri, C., Salarian, A., Carlson-Kuhta, P., Aminian, K., et al.: The instrumented timed up and go test: potential outcome measure for disease modifying therapies in Parkinson's disease. J. of Neurol., Neurosurg. & Psychiatry 81(2), 171–176 (2009)
5. Kohavi, R., John, G.H.: Wrappers for Feature Subset Selection. Art. Intel. 97(1-2), 273–324 (1997)