What are the optimal walking tests to assess disability progression?

Sébastien Piérard
Sebastien.Pierard@ulg.ac.be

Marc Van Droogenbroeck
M.VanDroogenbroeck@ulg.ac.be

INTELSIG, Montefiore Institute, University of Liège, Belgium

Introduction and objective

Therapy success is assumed when there is no evidence of disease activity, and especially no evidence of disability progression. Gait analysis plays a major role as gait impairment is considered by patients as the most disabling symptom [1]. Monitoring this phenomenon is thus important in the clinical setting. But how should we do this?

By selecting an appropriate technology, it is possible to measure many spatiotemporal gait parameters, describing both the swing and stance phases, even during long tests (*e.g.* 6min, 500m), without equipping patients with markers or sensors. Comfortable (at self selected speed), as fast as possible, and heel-to-toe (tandem) walking are typical tests. In this work, we determine if there is an advantage to perform various walking tests (in a reasonable amount of acquisition time), and which test or combination of tests is the most informative about the patient state. We focus on the ability to take objective decisions based on the measured gait parameters.

Method

- Many healthy volunteers (469 visits) and MS patients (86 visits) performed *12* tests:
  1. comfortable on a straight path of 25 ft (7.62 m)
  2. comfortable on a straight path of 25 ft (second try)
  3. as fast as possible on a straight path of 25 ft
  4. as fast as possible on a straight path of 25 ft (second try)
  5. tandem on a straight path of 25 ft
  6. tandem on a straight path of 25 ft (second try)
  7. comfortable on a ∞-shaped path of 20 m
  8. as fast as possible on a ∞-shaped path of 20 m
  9. tandem on a ∞-shaped path of 20 m
  10. comfortable on a ∞-shaped path of 100 m (5 laps of 20 m)
  11. as fast as possible on a ∞-shaped path of 100 m (5 laps of 20 m)
  12. as fast as possible on a ∞-shaped path of 500 m (25 laps of 20 m)
- All tests were recorded and analyzed with a gait measuring system *GAIMS* [2, 3, 4]. 26 gait parameters were computed for each test, and normalized with respect to those of healthy people with the same morphological characteristics (weight, height, age, gender, shoesize) as in [5].

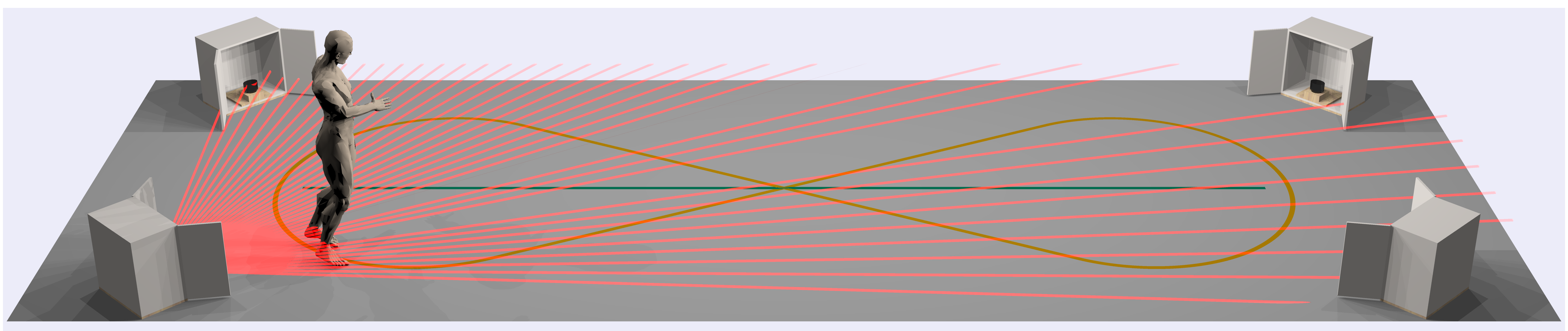
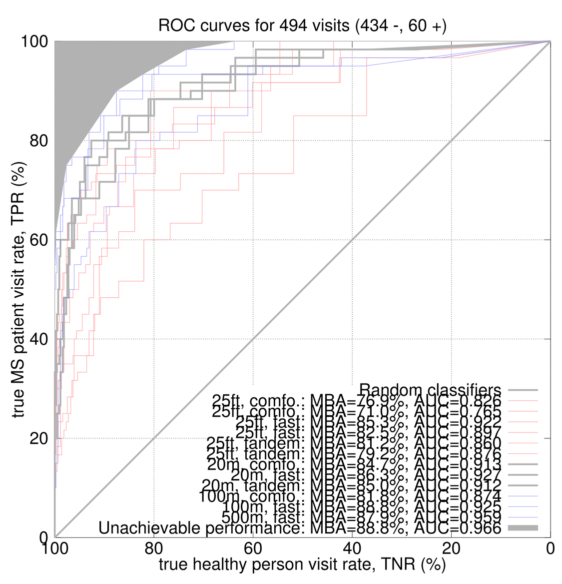


Figure 1: The gait measuring system *GAIMS* [2, 3, 4] measures feet trajectories with range laser scanners (the patient does not need to carry any marker or sensor) and derives many gait characteristics.

- To assess the ability to detect disability progression based on these clinical outcome measures, we evaluate the possibility of differentiating the people below a given EDSS threshold (0.75) from those above it based only on the measured gait parameters. We measure (by *leave-one-person-out*) the performance of the *12* classifiers learned automatically with the machine learning technique named *ExtRaTrees* (a forest of decision trees) [6]. They predict the probability of MS given the gait characteristics, and we apply a correction to express them with respect to a prior of $1/2$.



	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
T_1	1.00	0.53	0.46	0.44	0.38	0.34	0.49	0.47	0.34	0.48	0.47	0.46
T_2	0.53	1.00	0.40	0.44	0.44	0.39	0.48	0.42	0.35	0.51	0.40	0.37
T_3	0.46	0.40	1.00	0.66	0.37	0.34	0.54	0.63	0.37	0.51	0.59	0.61
T_4	0.44	0.44	0.66	1.00	0.34	0.33	0.49	0.64	0.37	0.49	0.58	0.60
T_5	0.38	0.44	0.37	0.34	1.00	0.65	0.40	0.43	0.64	0.43	0.46	0.44
T_6	0.34	0.39	0.34	0.33	0.65	1.00	0.32	0.42	0.60	0.37	0.41	0.40
T_7	0.49	0.48	0.54	0.49	0.40	0.32	1.00	0.52	0.38	0.61	0.47	0.49
T_8	0.47	0.42	0.63	0.64	0.43	0.42	0.52	1.00	0.48	0.53	0.73	0.67
T_9	0.34	0.35	0.37	0.37	0.64	0.60	0.38	0.48	1.00	0.44	0.48	0.46
T_{10}	0.48	0.51	0.51	0.49	0.43	0.37	0.61	0.53	0.44	1.00	0.57	0.58
T_{11}	0.47	0.40	0.59	0.58	0.46	0.41	0.47	0.73	0.48	0.57	1.00	0.73
T_{12}	0.46	0.37	0.61	0.60	0.44	0.40	0.49	0.67	0.46	0.58	0.73	1.00

Figure 2: The performance of the *12* classifiers (learned automatically from the corresponding gait tests) gives an indication about the quantity of useful information grabbed during the corresponding tests. Left: the ROC curves. Right: Pairwise Pearson correlations between the values predicted by different classifiers.

- As the classifiers are poorly correlated, we expect some benefits by combining them. Thus, we study the performance that can be achieved by combining the decisions of these *12* classifiers, or any subset of them. We assume that the optimal walking tests to assess disability progression are those leading to the best performance. Four combination strategies are analyzed: the probabilistic product [7], the median (related to a majority vote), the mean, and a weighted average (we determined the weights using a linear support vector machine with a soft margin and controlling its hyper-parameter C to avoid over-fitting and negative weights).

Results

The ability to detect the disability progression is quantified by the area under the ROC curve (AUC) and the maximum achievable balanced accuracy (MBA) of the combination. We show these metrics for each possible subset of tests, with respect to the total walking distance.

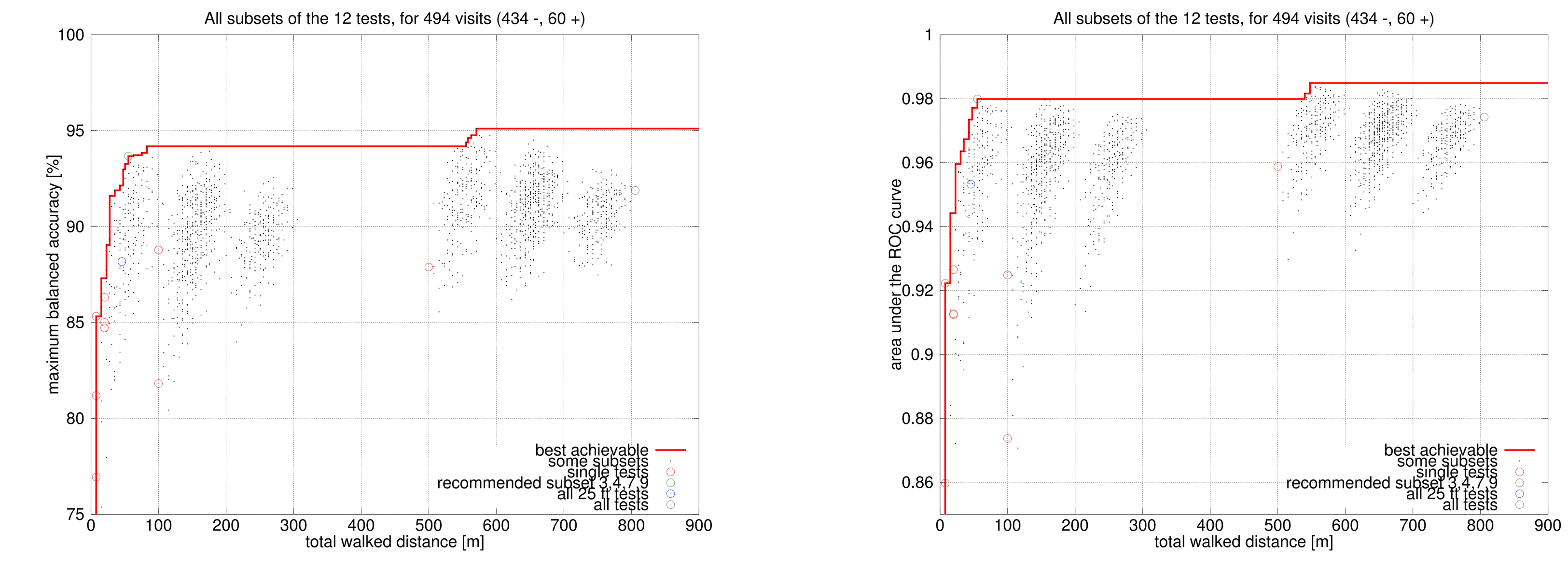


Figure 3: The ability to detect the disability progression depends on the set of gait analysis tests considered. This figure has been obtained with the combining rule "mean"; the behavior is similar for the other combination rules tested. Left: maximum balanced accuracy. Right: area under the ROC curve.

In order to choose the best set of clinical tests, we ranked them according to the two metrics, and for all four combination rules. Our advice is to select the set of tests that has the best (smallest) average rank. For the purpose of ranking, a small penalty of 1 % per 500 m has been considered for both the MBA and the AUC in order to prefer the shortest sets.

selected classifiers (walking tests)												total	ranks (/1727)								
T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}	walked	svm, w.r.t.		mean, w.r.t.		median, w.r.t.		product, w.r.t.		mean
												distance [m]	MBA	AUC	MBA	AUC	MBA	AUC	MBA	AUC	rank
		V	V			V		V				55.24	59	3	10	1	36	1	13	3	15.8
		V	V			V	V	V				75.24	53	13	3	3	12	41	3	2	16.2
		V	V	V		V		V				62.86	62	4	29	7	7	22	31	10	21.5
		V	V	V	V	V		V				70.48	9	1	67	19	11	13	65	1	23.2
		V	V	V	V	V						50.48	75	36	21	9	17	6	36	14	26.8
		V				V	V	V				67.62	71	26	56	12	27	18	2	4	27.0
		V	V	V	V	V	V	V				90.48	54	9	48	18	40	45	20	5	29.9
V		V	V			V	V	V				82.86	65	21	40	8	18	8	75	18	31.6
		V	V			V		V			V	555.24	17	78	16	37	34	53	6	15	32.0
V		V	V	V	V	V		V			V	578.10	23	69	35	57	24	47	19	27	37.6

Table 1: The top-10 sets of tests for clinical gait analysis. We recommend to rely on the tests 3, 4, 7, and 9. Considering other, less, or more tests would decrease the usefulness of clinical gait analysis.

An indicator about the patient's state can be obtained as follows. For each test in the selected set (3, 4, 7, and 9), the probability of MS is predicted based on the measures. These predictions are then combined by averaging. This gives us a correct decision rate (balanced accuracy) of 93.9 % when differentiating between $EDSS = 0$ (assumption for healthy people) and $EDSS \in [1.5, 5.5]$. But most importantly, it is possible to recognize most healthy people and most MS patients with 100 % certainty (the ROC curve follows the left and upper borders of the ROC space).

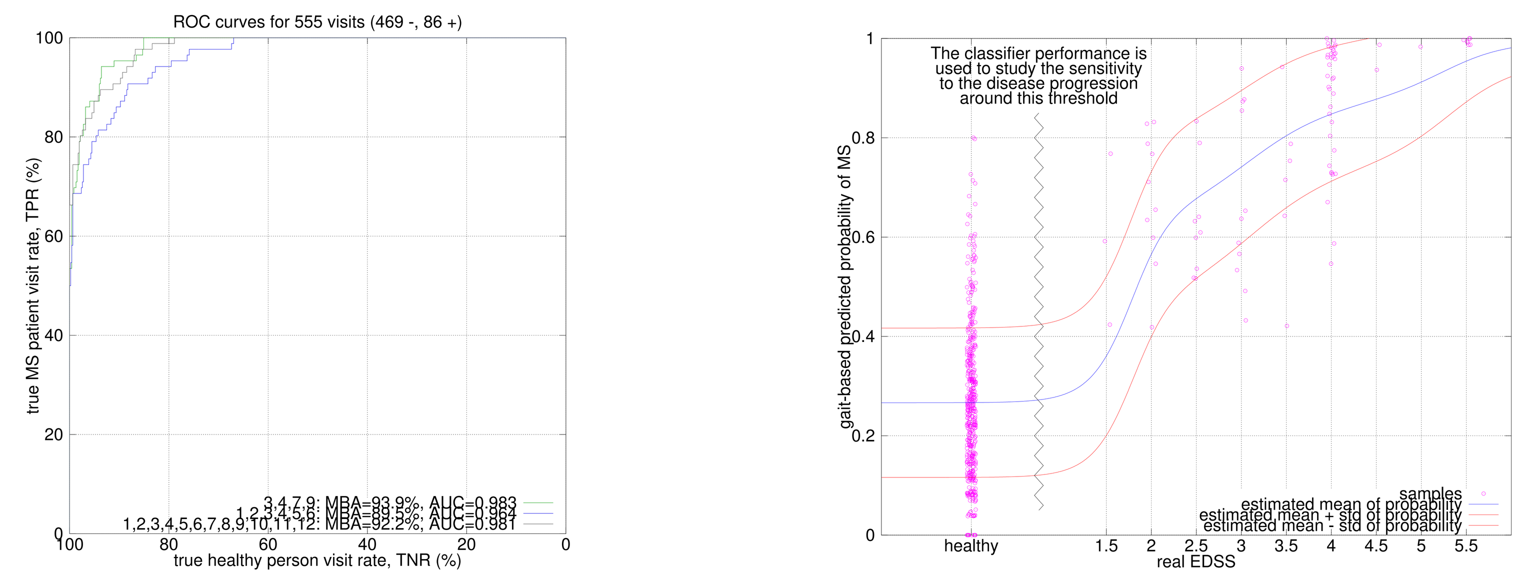


Figure 4: The usefulness of a gait-based predicted score (probability of MS) obtained by analyzing the measures taken during 4 tests, for a total walking distance of 55.24 m only. Left: the ROC curve corresponding to the ability to detect a modification of the EDSS around 0.75. Right: the relationship with the EDSS shows that the predicted probability of MS quickly increases in the first steps of the disease.

Conclusions

- With *an appropriate measuring system*, the clinical gait analysis can help to detect disability progression.
- Despite they are considered as standardized, *the shortest (25 ft) walking tests are the worst when considered alone*, while the longest (100 m, 500 m) ones are the best.
- Combining different walking tests improves the ability to take decisions, but the tests should be carefully selected. *Considering more tests than necessary can deteriorate the usefulness of clinical gait analysis.*
- We have studied 1727 sets of tests. *Our advice is to consider 4 tests for a total walking distance of 55.24 m only: two 25 ft walked as fast as possible, one 20 m with a comfortable pace on a ∞-shaped path, and one 20 m in tandem gait.*

References

[1] C. Heesen, J. Böhm, C. Reich, J. Kasper, M. Goebel, and S. Gold. Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. *Multiple Sclerosis*, 14:988–991, 2008.

[2] S. Piérard, R. Phan-Ba, V. Delvaux, P. Maquet, and M. Van Droogenbroeck. GAIMS: a powerful gait analysis system satisfying the constraints of clinical routine. In *Congress of the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS)*, Copenhagen, Denmark, October 2013.

[3] S. Piérard, S. Azrour, and M. Van Droogenbroeck. Design of a reliable processing pipeline for the non-intrusive measurement of feet trajectories with lasers. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4399–4403, Florence, Italy, May 2014.

[4] The GAIMS team. The GAIMS website. `http://www.montefiore.ulg.ac.be/gaims`, 2011.

[5] S. Azrour, S. Piérard, P. Geurts, and M. Van Droogenbroeck. Data normalization and supervised learning to assess the condition of patients with multiple sclerosis based on gait analysis. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 649–654, Bruges, Belgium, April 2014.

[6] P. Geurts, D. Ernst, and L. Wehenkel. Extremely randomized trees. *Machine Learning*, 63(1):3–42, April 2006.

[7] D. Tax, M. van Breukelen, R. Duin, and J. Kittler. Combining multiple classifiers by averaging or by multiplying? *Pattern Recognition*, 33(9):1475–1485, 2000.

Acknowledgements. We are grateful to the volunteers involved in this study, to the Walloon region of Belgium (`www.wallonie.be`) for partly funding it, and to BEA (`www.bea.be`) for the sensors.

Disclosure information. The authors have nothing to disclose.