S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.

Machine learning techniques to assess the performance of a gait analysis system

Sébastien Piérard, Rémy Phan-Ba, and Marc Van Droogenbroeck

INTELSIG Laboratory, Montefiore Institute, University of Liège, Belgium

Abstract. This paper presents a methodology based on machine learning techniques to assess the performance of a system measuring the trajectories of the lower limbs extremities for the follow-up of patients with multiple sclerosis. We show how we have established, with the help of machine learning, four important properties about this system: (1) an automated analysis of gait characteristics provides an improved analysis with respect to that of a human expert, (2) after learning, the gait characteristics provided by this system are valuable compared to measures taken by stopwatches, as used in the standardized tests, (3) the motion of the lower limbs extremities contains a lot of useful information about the gait, even if it is only a small part of the body motion, (4) a measurement system combined with a machine learning tool is sensitive to *intra-subject* modifications of the walking pattern.

1 Introduction

Although medical imaging is suited for the detection of severe neurological disorders, it is not always possible to diagnose a disease in its early stage. Therefore, as for patients with multiple sclerosis (MS) that often suffer from ambulation impairments, the clinical evaluation of the ambulation impairments is an appropriate method for the early detection of diseases. It is also useful for the follow-up of the patients. It helps to know, for example, if the proposed drugs and rehabilitation methods have a positive effect. However, the clinical evaluation is often limited in comparison to methods that rely on measuring technologies.

In the case of MS, clinicians use standardized tests mostly focusing on a distance or a velocity measurement. For example, the *timed-25 foot walk test* (T25FW) aims at measuring, with a stopwatch, the walking speed achievable by the patient on a $25\,ft$ straight path. But such tests have their limitations. For example, it has been established that if the speed evolution between two consecutive visits is less than 17% [1] or 20% [2], no relevant conclusion could be drawn on the evolution of MS.

Nevertheless, it has been proven, a long time ago by Gilles de la Tourette [3], that motion of the lower limbs extremities (named "feet" in the following) is a reliable source of information. More recently, Shao *et al.* [4], Teixidó *et al.* [5] and Piérard *et al.* [6] have proposed to rely on range laser scanners to measure the feet trajectories and to derive meaningful gait descriptors (GDs). Based on this idea, we have introduced in [7] a new gait measuring system, named *GAIMS*, designed to meet the constraints of the clinical routine.

Because of the complexity of gait analysis, it proved intractable to automate the diagnosis process made by a clinician. In addition, it was necessary to build

S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.

some evidence for the interest of a measuring systems such as *GAIMS*. The remainder of this paper discusses these issues, and hereafter we explain how machine learning was used to address the following four questions or issues:

[Q1] Can a gait analysis approach based on the objective measurements, provided by a system such as *GAIMS*, outperform the qualitative and subjective evaluation of a human expert?

[Q2] In [8], Phan-Ba *et al.* have established that at least two uncorrelated components are responsible for the *inter-subject* diversity of gait, for patients with MS. One component is the walking speed and the other one relates to the ataxia and gait asymmetry. But it was unknown which of them is the most powerful component for detecting the *intra-subject* gait changes over time. All in all, the question is whether *GAIMS* can outperform a simple stopwatch to detect gait abnormalities and their evolution.

[Q3] A system such as *GAIMS* only provides GDs related to the feet trajectories in a plane. At first sight, a lot of information seems to be missing: the knee angle, the foot orientation, the accelerations and forces, the minimum foot clearance (*i.e.* the minimum vertical distance between the foot and the floor during the swing phase), *etc.* However, those gait characteristics and the ones measured by *GAIMS* are not independent. Therefore, it is difficult to evaluate if *GAIMS* is capable of capturing enough information for the gait analysis of people with MS.

[Q4] In [9], Azrour *et al.* present some interesting findings related to the detection of MS, with an *ExtRaTrees* [10] classifier and the attributes derived from *GAIMS*. Such a study focuses on an *inter-subject* analysis since it does not address the gait evolution. But one of the key issues remains the capability to detect small *intra-subject* changes in the walking pattern of patients with gait disabilities.

In the following section, we present the experiments used to gather the data needed to address these four issues. Then, Section 3 discusses how the data was analyzed. Finally, in Section 4, we discuss our results and conclude the paper.

2 Experiments and data collection

As the disease evolves slowly with the progressive form of MS, it was impractical to involve patients in our study. We noted that cerebellar ataxia is known to be a major component of the gait impairments in MS, and that ataxia can be induced by alcohol. Therefore, we decided to rely on healthy persons with alcohol induced ataxia. Our experiment protocol has been approved by the ethics committee of the faculty of medicine of the university of Liège. Any reason (medical, physical, neurological, ethanol addiction, ...) that, in the opinion of the investigator, makes a subject unsuitable for enrollment has precluded his participation.

24 healthy volunteers, aged between 22 and 57 years, were asked to perform 23 tests, some before and others after alcohol intake, as shown in Table 1. In

S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.

	test	1	2	3	4	5	6	7	8	9	10	11	12	↑ alcohol intake and ↓ 30 minutes waiting	13	14	15	16	17	18	19	20	21	22	23
	alcohol														•	•	•	•	•	•	•	•	•	•	•
distance	$25\,ft$	•	•	•	•	•	•								•	•	•	•	•	•		•			
distance	$20\,m$							•	•	•											•	•	•		
distance	$100\,m$										•	•												•	•
distance	$500\,m$												•												
mode	comfortable	•	•					•			•				•	•					•			•	
mode	fast			•	•				•			•	•				•	•				•			•
mode	tandem					•	•			•									•	•			•		

Table 1: The 23 tests performed by the healthy volunteers. The data collected during the twelfth test are discarded in this study since no corresponding test has been performed after alcohol intake.

these tests, the volunteers were asked to walk in a given mode along a path drawn on the floor. Three paths were considered: a $25\,ft$ straight path, one turn of a ∞-shaped path of $20\,m$, and five turns of it. Three walking modes have been tested: comfortable (*i.e.* normal pace), as fast as possible, and *heel-to-toe* (*i.e. tandem gait*). The shortest tests are repeated twice to assess test-retest phenomena.

Between the twelfth and thirteenth tests, the participants were asked to consume ethanol (red wine, rum, or vodka). The blood alcohol concentration (BAC) was measured with a breathalyzer (Alert J5) thirty minutes after the end of acute ethanol ingestion (thirty minutes is the time required to reach the peak BAC). The volunteers were instructed to avoid eating and consuming "exciting" drinks (coca-cola, coffee, redbull, ...) 6 hours before the acquisition, and to avoid consuming alcohol 48 hours before. Note that we tried to reach the same peak BAC for all participants, and therefore the dose of alcohol administered has been determined with a normogram related to the gender and the weight. In practice, however, significant differences were observed between participants: the mean measured BAC is $67\,mg/l$ with standard deviation of $22\,mg/l$. At these concentrations, the most important modifications are behavioral, and the gait disorder specialists present during the acquisitions had difficulties to see any difference on feet movements induced by ethanol. Nevertheless, we show in the following that a difference exists, even if it is small (which was targeted). It should be noticed that several hypotheses could explain the difference (alcohol, motor fatigue, getting bored, *etc*), but its origin does not matter for us.

All tests were recorded with *GAIMS*, and the feet trajectories were computed with the processing pipeline presented in [11]. Based on these trajectories, 26 GDs were computed for each test. They are related to the speed, the inter-feet distance, the deviation from the followed path, the cadence, the stride length, the gait asymmetry, the temporal variability, and the proportion of double limb support time, to cite only a few.

3 Data analysis

In a preliminary step, we performed paired Student's t-tests in order to verify that a gait modification exists. It appears that many GDs significantly change after alcohol intake, but mostly for the $100\,m$ path in the preferred pace. We

S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.

found a decrease in the distance traveled ($p = 0.0044$), and the double limb support time ($p = 0.0155$), as well as an increase in the speed ($p = 0.0015$), the step length ($p = 0.00035$), the mean distance between feet ($p = 0.0002$), and the RMS value of the deviation from the followed path ($p = 0.00007$). However, these statistical results do not imply the possibility to detect the gait modification for each person. That is the purpose of the next type of analysis.

Let us consider a pair of tests (T_a, T_b), one being performed before alcohol intake, and the other one after alcohol intake. Two tests are paired if they correspond to the same person, to the same path, to the same walking mode, and if we do not mix test and retest. For each pair of tests, we try to recognize which test (first/second in the pair) was recorded after alcohol intake, based only on the GDs extracted by *GAIMS*. This comparison is expected to be easier than the prediction of a class (sober/alcohol) for each test. Indeed, the gait characteristics highly depend on the morphological characteristics, but we compare a person with himself in our approach. We adopt a binary classification in which the predicted class indicates whether the test led before alcohol intake is T_a or T_b. The set of attributes used to describe the pair of tests is given by

$$\omega\left(T_a\right), \pi\left(T_a\right), \left\{ \frac{f_i\left(T_a\right) + f_i\left(T_b\right)}{2}, f_i\left(T_b\right) - f_i\left(T_a\right), \frac{f_i\left(T_b\right) - f_i\left(T_a\right)}{f_i\left(T_a\right) + f_i\left(T_b\right)} \right\}_{i=1}^{26}$$

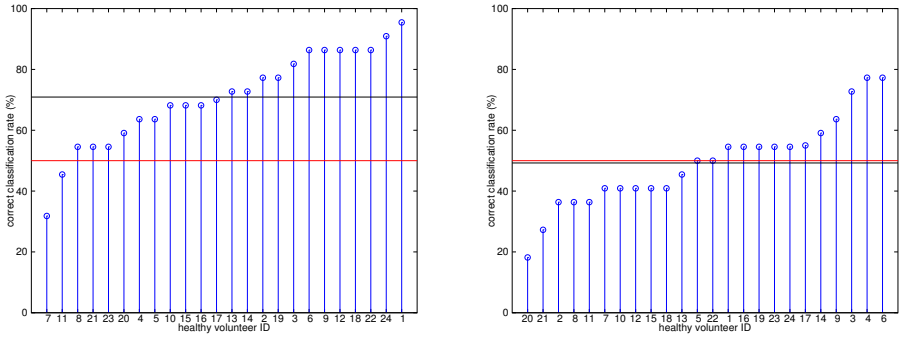
where $f_i\left(T\right)$ denotes the i-th GD of a test T, and $\omega\left(T\right)$ and $\pi\left(T\right)$ are symbolic attributes denoting the walking mode and the followed path respectively.

For the classification, we use a forest of randomized decision trees, named *ExtRaTrees* [10]. It does not require to optimize any parameter, nor to define a distance function, nor to select a kernel. It is insensitive to irrelevant attributes and can handle any relationship between them. But, as the *ExtRaTrees* are not expected to perform well with very small learning sets, we have considered the 11 types of tests all together, even if the GDs are not directly comparable for different tests. This multiplies by 11 the amount of learning samples. Moreover, our models gather 1000 trees to decrease the variance and to obtain reproducible results. The results reported hereafter have been obtained by a *leave-one-person-out* strategy: for each volunteer, his pairs of tests are classified using a model learned from the labeled pairs of tests of all other volunteers. To avoid the introduction of a bias, all the pairs are considered twice in the learning and test sets, once with the test recorded after alcohol intake in the first position and once in the last position. 24 models are thus learned from $(24-1)\times 11\times 2 = 506$ samples and tested on $1 \times 11 \times 2 = 22$ samples.

4 Results, discussion and conclusion

Based on all our 26 GDs, we obtain a correct decision rate of 70.9 % (see Figure 1a) on the pairs of tests. Considering multiple tests is advantageous since a majority vote permits to reach a correct decision rate of 91.7 % on the persons (for 22 volunteers out of 24, the majority of the tests are correctly recognized). This score demonstrates that our system is able to measure some interesting

S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.



(a) With all GDs, a machine learning based method is able to recognize the test recorded after alcohol intake most of the time: the overall correct decision rate is 70.9 % on the pairs of tests and 91.7 % on the persons after a majority vote.

(b) Using only the speed, a machine learning based method is unable to recognize the test recorded after alcohol intake: the overall correct decision rate is 49.2 %, which is symptomatic of a random guess. As expected, speed and ataxia are unrelated [8].

Figure 1: *GAIMS* produces an information superior to that of a stopwatch. This figure shows the correct decision rate w.r.t. the set of GDs considered: *GAIMS* on the left hand side, and a stopwatch on the right hand side. The vertical bars indicate the correct decision rate for the pairs of tests of each volunteer. The black horizontal line indicates the overall correct decision rate on the pairs of tests, which has to be compared with the correct decision rate that would be obtained with a random guess (*i.e.* 50 %, the red line).

information about the gait. The fact that it is often possible to recognize which test was recorded after alcohol intake indicates that it is possible to detect small *intra-subject* changes in the walking pattern of patients with gait disabilities such as those induced by MS *[answers to Q3 and Q4]*.

The score we have obtained was unexpected since the gait specialists present during the acquisition had difficulties to see any difference (only based on the feet movements) induced by alcohol intake. Furthermore, we have shown 228 randomly ordered pairs of video sequences (recorded during the acquisitions) to 14 other gait disorder specialists. They were able to correctly recognize the video sequence corresponding to the test performed after alcohol intake 142 times. Their correct decision rate (62.28 %, with a high inter-expert variability) is thus clearly lower than the one of the automatic classification system based on the GDs measured by *GAIMS* (70.9 %). Therefore, it turns out that an objective gait measurement system, such as *GAIMS*, is a valuable tool for helping neurologists to observe the walking difficulties encountered by their patients *[answer to Q1]*.

Finally, we want to establish if a gait analysis system combined to machine learning can outperform conclusions made with a stopwatch. Therefore, we have repeated the previous analysis, discarding the GDs that cannot be determined with a stopwatch. We obtained a correct decision rate of 49.2 % (see Figure 1b)

S. PIÉRARD, R. PHAN-BA, and M. VAN DROOGENBROECK. **Machine learning techniques to assess the performance of a gait analysis system**. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, pages 419-424, April 2014.

on the pairs of tests. The system is thus unable to find any relationship between the remaining attributes and the gait modification. We therefore conclude that *GAIMS* is appropriate for detecting cerebellar ataxia *[answer to Q2]*.

Even if our results are encouraging, it should be noted that the small size of our database is a limitation for the learned models. Our results could certainly be improved with supplementary volunteers. Moreover, we did not try to tune the set of attributes, and we did not try various kinds of classifiers yet.

In future works, we will enhance the data set, but we have already established that machine learning tools are appropriate for the interpretation of gait data. This would otherwise be out of reach for the clinical diagnosis.

Acknowledgements

We are grateful to the volunteers involved in this study, to Shibeshih Belachew and Gustave Moonen for their valuable support, to the Walloon region of Belgium (`www.wallonie.be`) for partly funding this project, and to BEA (`www.bea.be`) for the sensors.

References

[1] C. Coleman, D. Sobieraj, and L. Marinucci. Minimally important clinical difference of the Timed 25-Foot Walk Test: results from a randomized controlled trial in patients with multiple sclerosis. *Current Medical Research and Opinion*, 28(1):49–56, January 2012.

[2] S. Schwid, A. Goodman, M. McDermott, C. Bever, and S. Cook. Quantitative functional measures in MS: What is a reliable change? *Neurology*, 58(8):1294–1296, April 2002.

[3] H. Acharya and R. Camicioli. The forgotten thesis of Gilles de la Tourette. *University of Alberta Health Sciences Journal*, 2(2):32–33, December 2005.

[4] X. Shao, H. Zhao, K. Nakamura, R. Shibasaki, R. Zhang, and Z. Liu. Analyzing pedestrians' walking patterns using single-row laser range scanners. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 2, pages 1202–1207, Taipei, Taiwan, October 2006.

[5] M. Teixidó, T. Pallejà, M. Tresanchez, M. Norgués, and J. Palacín. Measuring oscillating walking paths with a LIDAR. *Sensors*, 11:5071–5086, 2011.

[6] S. Piérard, M. Van Droogenbroeck, R. Phan-Ba, and S. Belachew. A new low-cost and non-intrusive feet tracker. In *Workshop on Circuits, Systems and Signal Processing (ProRISC)*, pages 382–387, Veldhoven, The Netherlands, November 2011.

[7] 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. *Multiple Sclerosis Journal*, 19(S1):359, October 2013. Proceedings of ECTRIMS/RIMS 2013 (Copenhagen, Denmark), P800.

[8] R. Phan-Ba, S. Piérard, E. Lommers, G. Delrue, P. Calay, M. Van Droogenbroeck, V. Delvaux, and P. Maquet. Relative contribution of walking speed, ataxia and gait asymmetry to the composition of gait in multiple sclerosis. *Multiple Sclerosis Journal*, 19(S1):261, October 2013. Proceedings of ECTRIMS/RIMS 2013 (Copenhagen, Denmark), P608.

[9] S. Azrour, S. Piérard, and M. Van Droogenbroeck. Using gait measuring system (GAIMS) to discriminate patients with multiple sclerosis from healthy persons. In *BEMEKO Workshop on measurement: Challenges and Opportunities*, Liège, Belgium, November 2013.

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

[11] 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)*, Florence, Italy, May 2014.