ACCELEROMETER BASED GAIT ANALYSIS
Multi Variate Assessment of Fall Risk with FD-NEAT

Bart Jansen[1], Maxine Tan[1], Ivan Bautmans[2,3,4,5], Bart Van Keymolen[2,3]
Tony Mets[2,3,4] and Rudi Deklerck[1]

[1] Departement Electronics and Informatic and IBBT, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2] Gerontology, Frailty in Ageing Research Department, Vrije Universiteit Brussel
Laarbeeklaan 103, 1090 Brussels, Belgium
[3] Frailty in Ageing Research Department, Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussels, Belgium
[4] Geriatrics, Universitair Ziekenhuis Brussel, Laarbeeklaan 101, 1090 Brussels, Belgium
[5] Stichting Opleiding Musculoskeletale Therapie (SOMT), Softwareweg 5, 3821 BN Amersfoort, The Netherlands

Keywords: Accelerometry, Fall risk, Gait analysis, Step time asymmetry, Classification, Feature selection, FD-NEAT.

Abstract: This paper describes an accelerometer based gait analysis system for the assessment of fall risk. The assessment is based on 22 different features calculated from the signal. The different features are combined using machine learning algorithms in order to decide whether the subject has an increased fall risk. Results from Naive Bayes, Neural Networks, Locally Weighted Learning, Support Vector Machines and C4.5 are reported and compared. It is argued that the neural networks provide low accuracy results because of the high dimensionality of the feature space compared to the available data. It is shown that FD-NEAT (a method from neuro evolution which simultaneously learns the network topology, the network weights and the relevant features) outperforms the other methods in the given classification task. The system is evaluated on a database consisting of 40 elderly with known fall risk and 40 healthy elderly controls.

1 INTRODUCTION

The field of accelerometer based fall risk assessment is characterised by an intense debate on the relevance of some specific features calculated from the accelerometer signal in a univariate classification problem on the distinction between fallers from non fallers (e.g. Moe-Nilssen and Helbostad, 2005).

Many of these features are known for decades (e.g. a decrease in step length is related to an increase in fall risk); others could only be described since the availability of small, battery powered accelerometer sensors (e.g. step regularity and symmetry). Only recently, the validity, reliability and repeatability of most of these features in the context of fall risk assessment have been described in the clinical literature (e.g. Moe-Nilssen, 1998).

For some specific diseases and conditions having a direct impact on the gait pattern, it is well described how the disease is affecting the neurological or muscolatory system and how this affects the accelerometer based gait features. For example, in Parkinson patients, the freezing of gait increases the variability in stride time and the effect of specific treatment on the freezing of gait can be evaluated by investigating stride time variability (Hausdorff et al, 2005).

However, in a majority of the growing population of elderly, an increase in fall risk cannot directly be attributed to a specific disease. Rather, a condition of general frailty, multiple chronic diseases and a general decrease in mobility all together contribute to the increased fall risk. Therefore, it is to be expected that in a general population of elderly, a less clear relationship between single accelerometer based gait features and fall risk can be observed. However, very few attention was paid so far to the construction of intelligent multi-variate classifiers for fall risk assessment.

This paper evaluates the use of the FD-NEAT algorithm (Tan et al, 2009) for the classification of a population of eighty elderly into a class of elderly presenting increased fall-risk and a class of elderly without an increased risk of falling, based on a wide variety of accelerometer based gait features.

From a machine learning perspective, the problem is far from trivial, as there is extremely few training data available, compared to the large number of features considered. We will show that FD-NEAT outperforms traditional neural networks and other machine learning algorithms, because it better copes with the dimensionality problem, by means of intelligent feature selection. Because of the clinically unclear relationship between many of the features used in this paper and fall risk in a general population of elderly, no a priori feature selection was performed based on the available medical knowledge.

2 PREVIOUS WORK

2.1 Accelerometer based Fall Risk Assessment

Most studies in accelerometer based gait analysis for fall risk assessment are focusing on the repeatability or validity of single outcomes. Only very few studies are combining multiple outcomes in intelligent classifiers. Recently, Marschollek et al. showed how measures obtained from accelerometry could be combined with clinical scores, in order to discriminate between fallers and non-fallers during an instrumented timed up and go (TUG) test (Marschollek et al., 2009). Another study, combining results from accelerometer based TUG tests, stepping tests and a sit-to-stand tests, is reported by (Narayanan, 2010).

(Swanenburg et al, 2010) report on a one year prospective study of fall risk assessment, based on features calculated from force plates. Intelligent classifiers based on neural networks were also used for fall risk assessment from posturography tasks, instrumented with accelerometer and gyroscopes (Giansanti et al. 2008).

3 METHODS

3.1 Subjects

Eighty subjects participated in this study. They consisted of two groups (n=40 each): elderly with known fall risk (EF) and elderly controls (EC), see table 1. Each group contained twenty males and twenty females.

Table 1: General subject information. None of the parameters was significantly different between the groups (ANOVA $p < 0.05$). EF = elderly with increased fall risk, EC = elderly controls. Standard deviations are shown between brackets.

	Group	
	EF (n=40)	EC (n=40)
Age	80.59 (5.38)	79.03 (4.95)
Weight	66.89 (14.97)	69.74 (11.56)
Length	1.62 (0.12)	1.64 (0.08)
BMI	25.46 (4.29)	25.61 (3.93)

All subjects were older than 70. Known fall risk was defined as a reported history of falls and/or Timed-Get-Up-and-Go-Test > 15s and/or Tinetti test $\leq 24/28$. The local ethical committee approved this study and all participants provided their written informed consent.

3.2 Data Acquisition

The DynaPort Minimod tri-axial accelerometer (McRoberts BV, The Hague, The Netherlands) was placed at the sacrum of the subjects. The device stores the accelerometer signal on a standard SD-card. Before every walking episode, the SD-card was emptied, put in the sensor and the sensor was restarted. After every walking episode, the SD-card was placed in the laptop and the acceleration data along the three axes was read out using the Mira software (same manufacturer) and exported to plain text files. The plain text files were loaded into our own gait analysis toolbox, an in-house developed software package programmed in C#.

3.3 Test Procedure

Subjects were asked to walk a straight line trajectory of 18 meters, separated by two clear lines on the floor. Subjects started with both feet in front of the first line and stopped when the second foot landed beyond the second line. The distance between the stop line and the final heel strike was measured and added to the 18 meter to obtain the total distance walked. At the beginning of each walk, the observer placed the sensor at the sacrum and initialized the sensor as described above. Subjects were always instructed to walk at preferred speed and no walking aids were allowed.

3.4 Statistical Analysis

A dataset of eighty elderly consisting of forty elderly with increased fall risk and forty controls is available. Compared to similar datasets used in

accelerometer based gait analysis, this is quite a large set. However, from a machine learning perspective given the high amount of included features, its size is extremely small. Therefore, it was decided not to partition the dataset into a single training set and a single test set, but ten-fold cross validation was used (Duda et al, 2000).

Data analysis was performed using SPSS version 17, WEKA and Excel. For each of the studied machine learning algorithms averages over the ten folds of accuracy, true positives, false positives, precision, recall and Area under the Curve (AUC) are provided.

Significant differences among outcome measures are evaluated using a one way ANOVA with significance set to $p<0.05$ and with a post-hoc Bonferroni test to identify two differing measures. Correlations between different types of features were assessed by calculating Pearson's correlation coefficient.

3.5 Data Analysis

In total, 22 features were calculated from the accelerometer signal. Features can be divided into five groups: step count, step time (and derived statistics), step length (and derived statistics), step symmetry and step RMS. Each of the five groups is explained below.

3.5.1 Step Count

The 3D accelerometer signal is rotated to align the Y axis of the signal to gravity and steps (defined as initial contacts of the heel (IC)) are identified, based on the maxima before the zero-crossings in the forward acceleration signal, after applying a fourth order zero lag Butterworth low pass filter with a cut-off frequency of 2 Hz (Zijlstra, 2004).

3.5.2 Step Time

From the IC's detected from the signals as described above, the average step time is obtained, as well as a range of derived statistics including standard deviation, coefficient of variation, inter quartile range etc. Also, step frequency, walking speed and step time asymmetry are available. Step time asymmetry is the difference of the left step time and the right step time, scaled by the average and expressed as a percentage (equation 2).

$$t_{asym} = 200 \frac{|t_{left} - t_{right}|}{t_{left} + t_{right}} \qquad (2)$$

In all features based on step time, the initial two steps and the final two steps are discarded from the signal, in order to exclude effects from gait initiation and gait termination. The study of irregularities in the gait initiation and termination phases is beyond the scope of this paper.

3.5.3 Step Length

As the total length of the trajectory and the number of steps are known, the average step length is available. Step length can also be calculated without relying on the measured true trajectory length using the inverted pendulum model (Zijlstra, 2004). The inverted pendulum model is a biomechanical model of human gait, which is relating a vertical movement of the pelvis during the gait cycle with the step length, as specified in equation 3:

$$steplength = 2\sqrt{2lh - h^2} \qquad (3)$$

where l is the leg length and h is the vertical displacement of the pelvis, which is obtained from the double integration of the vertical acceleration component. The advantage of this method is that the step length of each gait cycle can be calculated individually, using only two parameters. The disadvantage of this method is that the double integration step in calculating the vertical displacement is prone to drift. Step length as calculated from the known trajectory length, step length according to the inverted pendulum model and the vertical displacement of the pelvis itself are included as features of this study.

3.5.4 Step and Stride
Regularity and Symmetry

A whole family of related measures exist which all capture the regularity of the accelerometer signal over multiple steps (Moe-Nilssen and Helbostadt, 2004). Suppose y[t] is the auto correlation of the acceleration signal a[t], then y[t] has maxima corresponding to a time shift of 1,2,.. k steps. Hence, the auto correlation at the first maximum expresses the step regularity, whereas auto-correlation at the second maximum expresses the stride regularity. Step symmetry is defined as the step regularity divided by the stride regularity.

These measures are typically calculated in the medio-lateral (ML) and cranio-caudal (CC) direction. Auto correlation could be normalized or not. Biased and unbiased versions have been proposed. In this study only the unbiased measures in the CC orientation were incorporated.

3.5.5 Step and Stride RMS

The root mean squared acceleration per step or per stride, in the CC and the ML direction were calculated (equation 4).

$$RMS = \sqrt{\frac{\sum a_i{}^2}{N}} \qquad (4)$$

a_i is the acceleration in the i-th sample of the considered step in either the CC or ML orientation and N is the number of samples in the current step. RMS values per step are averaged over all steps in the signal.

3.6 Machine Learning Algorithms and Classifiers

Standard machine learning methods for pattern recognition and classification were employed: Naive Bayes (NB), Multi layered perceptron (MLP), Support Vector Machines (SVM), Locally Weighted Learning (LWL) and C45. For more information on any of these classifiers, the reader is referred to (Duda et al, 2000). WEKA version 3.4.13 was used for the classification based on each of these classifiers. Weka is an open source machine learning software package, provided by the University of Waikato. For each of the included algorithms, the default parameters as proposed by WEKA were used.

Naive Bayes. This is a simple Bayesian classifier, assuming conditional independence between the attributes.

Multi layered Perceptron. We have used a MLP with one hidden layer, consisting of 22 input units, 12 hidden units, which is (nr inputs + output values) / 2, and a single output unit. Learning rate was set to 0.3 and momentum to 0.2. In the hidden units a sigmoid activation function was used, in the output unit the identity function was used as an activation function. Back propagation was used as a training algorithm, performing 500 iterations.

Support Vector Machines. Sequential Minimal Optimization is used for fast training (Platt J, 1998).

Locally Weighted Learning. This is a nearest neighbor classifier, which is considering all neighbors and attributing a weight to each of the neighbors. The weight scales linear with the inverse distance to the query point (Atkeson, Moore and Schaal, 1996).

C45. is a traditional decision tree learning algorithm introduced by (Quinlan, 1993).

3.7 FD-NEAT

Training neural networks for classifying gait data with back propagation as used in the MLP, has several drawbacks: (1) the user must define the network topology; (2) the user must carefully select the relevant features, as (3) a fastly growing number of training instances is required with the addition of each new attribute. In the current study, only 80 subjects are available, which might hence lead to low classification accuracy of neural networks.

Approaches based on genetic algorithms were proposed in which the network topology and the weights are learnt simultaneously. An example of such a system is "neuro evolution of augmenting topologies" or NEAT (Stanley and Miikkulainen, 2002) in which a population of neural networks evolves from simple perceptrons into more complex networks, based on mutation (both weights and connections evolve) and cross-over. NEAT was shown to perform superior to classical neural nets in typical benchmark problems. A straightforward extension of NEAT is "feature selective" NEAT or FS-NEAT (Whiteson et al, 2005) which performs feature selection, topology learning and weight learning simultaneously. In FS-NEAT the initial networks in the population only have a single input neuron, randomly selected from the available attributes. A mutation operator which can connect additional input nodes is added. A modification to FS-NEAT is "feature deselective" NEAT or FD-NEAT (Tan et al, 2009), which is similar but starts from networks in which all inputs are connected and has a mutation operator which drops connections from input nodes. In most cases, FD-NEAT outperforms FS-NEAT.

In the experiments we report in this article, the best result out of ten runs was obtained for each fold. The populations in the genetic algorithm consisted of 200 networks that were evolved over 60 generations.

4 RESULTS

4.1 Comparison of Classifiers

The main experiment consists of evaluating the performance of five different machine learning algorithms in a binary classification problem based on 22 features calculated from an accelerometer signal obtained from a single walk of 18 m. Detailed results are given in table 3. In summary, the results show that NB outperforms the other classifiers: it

has the highest accuracy (0.77), true positive rate (= recall) (0.75), precision (0.79) and area under the curve (0.82), for the lowest false positive rate (0.2). The multi layered perceptron scored very low, on each of the five performance measures incorporated in this study. Given the high amount of attributes (22) compared to the low amount of training instances (72 out of 80 in each fold), it is to be expected that inferior results of the MLP are due to a severe lack of training instances compared to the complexity of the network. Feature selection is hence appropriate.

Table 2: Results of the classifiers, 22 features. TP = true positives, FP = false positives, AUC = area under the curve, NB = Naive Bayes, MLP=Multi Layered Perceptron, SVM = Support Vector Machine, LWL = Locally Weighted Learning. X= AUC for FD-NEAT not available, see text.

	accuracy	TP	FP	precision	recall	AUC
NB-22	.77	.75	.2	.79	.75	.82
MLP-22	.61	.6	.38	.62	.6	.72
SVM-22	.69	.6	.23	.73	.6	.69
LWL-22	.69	.58	.2	.74	.58	.74
C45-22	.69	.65	.28	.70	.65	.64
FD-NEAT	.82	.8	.15	.84	.8	X

Given the inconclusive results of the debate in the gait analysis community on the relevance of each of the individual features, it is not advised to manually select the relevant features. On the other hand, one of the main observations in the field of clinical gait analysis is that almost all features are somehow influenced by walking speed. In this study, subjects were asked to walk at normal speed. Hence, differences in any of the calculated features between EF and EC may be related to gait speed differences between both groups.

Using FD-NEAT the accuracy increases up to 82.5%. Also TP (recall) and precision are the highest of all experiments reported, while FP is the lowest of all reported experiments. For FD-NEAT ROC analysis with AUC could not be reported, as it uses a sigmoid in the activation function, resulting in nearly binary outputs such that threshold varying is unfeasible.

Standard deviations of the accuracy over the ten folds were calculated for MLP-22 and FD-NEAT-22 and are quite high (σ=0.17 and 0.17 respectively). This is caused by the too small fold size (for N=80, the fold size is 8). At the 0.05 level,

accuracy of FD-NEAT-22 is significantly better than MLP-22.

5 DISCUSSION

From a clinical perspective, this study confirms that accelerometer based fall risk assessment is feasible with high accuracy (82.5%) and with high sensitivity (80% recall). However, the study population was recruited based on self reported falls, the timed up and go test and the Tinetti test. Hence, amongst EF, a wide variety of conditions and diseases which are possible related to fall risk are present. In a study design in which only fallers suffering from a specific disease or condition are included (e.g. sarcopenia or Alzheimer's disease), higher accuracy results could probably be obtained, using another subset of features, as each disease results in specific gait disorders. As this article is focussing on the screening potential of accelerometer based gait analysis for fall risk, we have chosen not to restrict the study population to specific subgroups.

Most measures employed do not show significant differences between the different classifiers studied. This is due to the high standard deviations as the sizes of the folds studied are extremely small. However, validating over the test set was not considered a viable approach. Hence, it is to be advised to repeat the experiment over larger population sizes in order to reach significance.

Nevertheless, FD-NEAT based on 22 features significantly outperforms MLP based on 22 features, confirming our initial hypothesis that FD-NEAT suffers less from the described dimensionality problems.

6 CONCLUSIONS

This article evaluated the possibility of fall risk stratification of elderly based on a single walk of 18 meters, instrumented with an accelerometer. Opposed to many systems, the system is not limited to a single feature. We've investigated the performance of five different classifiers using 22 features, commonly used in various gait experiments. Given the extremely small data set (40 positive and 40 negative cases) compared to the number of attributes (22), the performance of the classifiers is suboptimal (60 to 70 % accuracy). We've put forward that FD-NEAT, an evolutionary approach to perform feature selection, to learn a

neural network topology and to learn the weights simultaneously, outperforms the traditional classifiers (82.5 % accuracy).

From a clinical perspective, this article illustrates that in a general population of elderly, fall risk is related to different underlying constructs, with clear manifestations among different dimensions in the gait pattern as captured by the accelerometer.

REFERENCES

Atkeson C., Moore A., and Schaal S., 1996. "Locally weighted learning" AI Reviews.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*. Wiley-Interscience.

Giansanti, D., Maccioni, G., Cesarino, S., Benvenuti, F., & Maccellari, V. (2008). Assessment of fall-risk by means of a neural network based on parameters assessed by a wearable device during posturography. *Medical Engineering and Physics*, *30*, 367-372.

Hausdorff, J. M., Schaafsma, J. D., Balash, Y., Bartels, a. L., Gurevich, T., Giladi, N., et al. (2003). Impaired regulation of stride variability in Parkinson's disease subjects with freezing of gait. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *149*(2), 187-94.

Hausdorff J. M., 2005. Gait variability: methods, modeling and meaning. *Journal of NeuroEngineering and Rehabilitation*. 9:1-9.

Marschollek, M., Nemitz, G., Gietzelt, M., Wolf, K. H., Meyer Zu Schwabedissen, H., Haux, R., et al., 2009. Predicting in-patient falls in a geriatric clinic: a clinical study combining assessment data and simple sensory gait measurements. *Zeitschrift für Gerontologie und Geriatrie : Organ der Deutschen Gesellschaft für Gerontologie und Geriatrie*, *42*(4), 317-21.

Marschollek M., Wolf K., Plischke M., Haux R., 2006. Classification of activities of daily life from long-term realistic multi-sensor data. In: *Proceedings of IEEE Health Pervasive Systems HPS06*.

Moe-Nilssen, R. (1998). Test-retest reliability of trunk accelerometry during standing and walking *1, *2. *Archives of Physical Medicine and Rehabilitation*, *79*(11), 1377-1385.

Moe-Nilssen R., Helbostad J. L., 2004. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics*. 37(1):121-126.

Moe-Nilssen, R., & Helbostad, J. L. (2005). Interstride trunk acceleration variability but not step width variability can differentiate between fit and frail older adults. *Gait & posture*, *21*(2), 164-70.

Narayanan, M. R., Redmond, S. J., Scalzi, M. E., Lord, S. R., Celler, B. G., Lovell Ast, N. H., et al., 2010. Longitudinal falls-risk estimation using triaxial accelerometry. *IEEE transactions on bio-medical engineering*, *57*(3), 534-41.

Oshima Y., Kawaguchi K., Tanaka S., et al., 2010. Classifying household and locomotive activities using a triaxial accelerometer. *Gait & posture*. 31(3):370-4.

Platt J., 1998. "Fast Training of Support Vector Machines using Sequential Minimal Optimization". Advances in Kernel Methods-Support Vector Learning, B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press.

Quinlan R., 1993. "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, San Mateo.

Stanley K. O., Miikkulainen R., 2002. Evolving neural networks through augmenting topologies. Evol Comput 10:99–127.

Swanenburg, J., de Bruin, E. D., Uebelhart, D., & Mulder, T. (2010). Falls prediction in elderly people: a 1-year prospective study. *Gait & posture*, *31*(3), 317-21.

Tan M., Hartley M., Bister M., Deklerck R., 2009. Automated feature selection in neuroevolution. *Evolutionary Intelligence*. 1(4):271-292.

Whiteson S., Stanley K. O., Miikkulainen R., 2004. Automatic feature selection in neuroevolution. In: Proceedings of the genetic and evolutionary computation conference (GECCO), Seattle, Washington, USA.

Zijlstra W., 2004. Assessment of spatio-temporal parameters during unconstrained walking. *European journal of applied physiology*. 92(1-2):39-44.