

Original scientific paper

A DATA-MINING BASED METHOD FOR THE GAIT PATTERN ANALYSIS

UDC 004.8:617.3

**Marcelo Rudek¹, Nicoli Mariá Silva¹,
Jean-Paul Steinmetz², Andreas Jähnen³**

¹Pontifical Catholic University of Parana, Department of Production and System Engineering, Brazil

²Department of Research and Development, Zitha Senior - Centre on Memory and Mobility Michel Rodange, Zitha, Luxembourg

³Luxembourg Institute of Science and Technology (LIST), Luxembourg

Abstract. *The paper presents a method developed for the gait classification based on the analysis of the trajectory of the pressure centres (CoP) extracted from the contact points of the feet with the ground during walking. The data acquirement is performed by means of a walkway with embedded tactile sensors. The proposed method includes capturing procedures, standardization of data, creation of an organized repository (data warehouse), and development of a process mining. A graphical analysis is applied to looking at the footprint signature patterns. The aim is to obtain a visual interpretation of the grouping by situating it into the normal walking patterns or deviations associated with an individual way of walking. The method consists of data classification automation which divides them into healthy and non-healthy subjects in order to assist in rehabilitation treatments for the people with related mobility problems.*

Key words: *Gait Analysis, Pattern Analysis, Data Mining, Rehabilitation.*

I. INTRODUCTION

Actual monitoring technologies based on biometric signals have allowed important improvements in healthcare applications and assistive uses. The approaches that include biomechanics analysis and its functional signaling which focus on rehabilitation or early disease detection are being investigated in many ways as in [1-5, 19, 20]. New systems using cameras, forceplates, walkways and insole devices offer new perspectives to data

Received October 10, 2015 / Accepted November 10, 2015

Corresponding author: Marcelo Rudek

Pontifical Catholic University of Parana, PUCPR/ Production and System Engineering Graduate Program – PPGEPS, Imaculada conceicao, 1155, 80215-030, Curitiba, Brazil

E-mail: marcelo.rudek@pucpr.br

analysis as in [6]. Nowadays an increasing field is observed with the use of sensors to achieve informative data applied to the evaluation of movement of a person. The tactile sensors generate pressure distribution maps from the feet contact during walking [7]. This mapping might be used in the measurements of posture, equilibrium and motor control of the patients within the clinical evaluation of the walking process. Measuring stability allows a better understanding of the people's necessities.

In this context, the research demonstrates a special interest in the investigation of elderly people [8]. The goal is to provide support to the patients susceptible to the risk of falls due to the deterioration of their mobility caused by neurological disorders, diseases or surgery that affect their motor control [9]. For example, the systems can provide a data analysis during the patient treatment period in his physical rehabilitation to evaluate his self-control in the movements' progression. Four essential features about the gait analysis are described in [9]. The first is how much and in what direction the given patient leans while standing still. The second feature is how much pressure the patient exerts under each foot and in what regions of the foot. The third clinical feature is how much the patient sways while attempting to stand still. The fourth clinical feature characterizes dynamic balance. In motion the whole set of features occurs at same time; that is why it is not possible to fully capture the essence of motion without tactile information [10].

There are several applications in which the medical professional needs information of healthy subjects that can be used as reference for comparison with the patient's behavior as in [6, 12, 13]. The pressure maps produce this relationship. One possible way of getting the visualization of the pressure map of walking is by the trajectory of the centre of pressure (CoP) under each footprint in the walking map. In the standing position, as the method of [10, 11], it is used a sensor plate system to generate a graphical representation of the CoP variation determined by averaging the pixel values on the location of the intensity-weighted pressure points of each foot. In that case, the CoP changes according the equilibrium for the overall image, which could be considered an estimate of the person's centre of gravity (CoG). As proved by [1, 8, 9, 10, 14 and 15], it is possible to estimate the balance information from CoP. Then, the CoP analysis is very useful to understand the gait pattern of an individual for his rehabilitation evaluation.

The information used in this research is from a tactile sensor mattress (walkway) [7]. This device generates a map containing the pressure variations of the contact points of foot during walking. This enables us to conduct the study by achieving the walking characteristics of each individual, i. e.; it provides a way to carry out balance measurement and cadence motor control analysis of each individual. The special interest is a gait analysis in elderly patients with a potential risk of falling; thus, it can act to prevent the risks and build support devices customized to the individual walking characteristics.

In this wide context, the main objective of the research is to define a separation process using the automated data mining of the behavior pattern by classifying healthy or unhealthy people, or classifying different pathologies that interfere with the movement.

2. PROPOSED METHOD

The functional structure of proposed method for the gait pattern classification is shown in Fig. 1. Basically, the process model has two main parts; a pool dedicated to the "Data Gathering and Analysis" (i.e. clinical viewpoint) and the "Data Processing" contains four

inner process, as i) Data Normalization; ii) Data Warehouse; iii) Data Mining and iv) Graphical Representation. These steps are discussed in the following sections as well as the background concepts about data acquirement and their respective parameters.

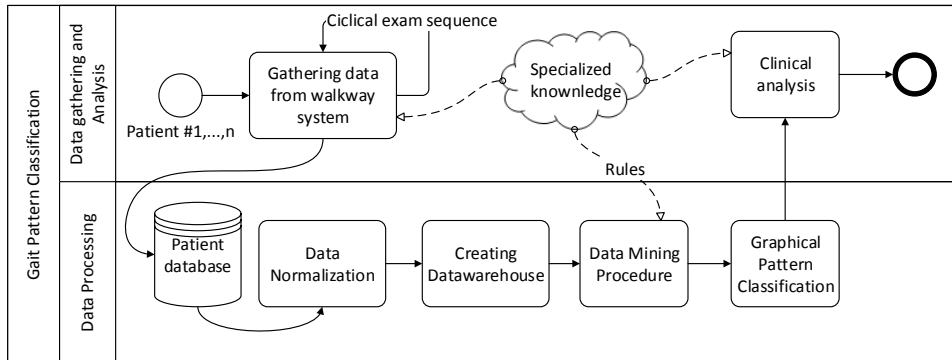


Fig. 1 Process modelling of the gait pattern classification

2.1. Gait data background

The Gaitrite® [7] system is a complete tool for measuring the gait information to the indoor laboratorial approach. During the patient movement while crossing the walkway, the pressure points affected by the footfall generate two groups of measured values called spatial and temporal parameters. Spatial Parameters are the step length, stride length, base width, step width, stride width, toe in / toe out, leg length, length of foot, and width of foot. The temporal parameters are step time, stride time, stance time, swing time, ambulation time, velocity, cadence, single support, double support. Fig.2 shows the spatial parameters used in our study and definitions are in [7].

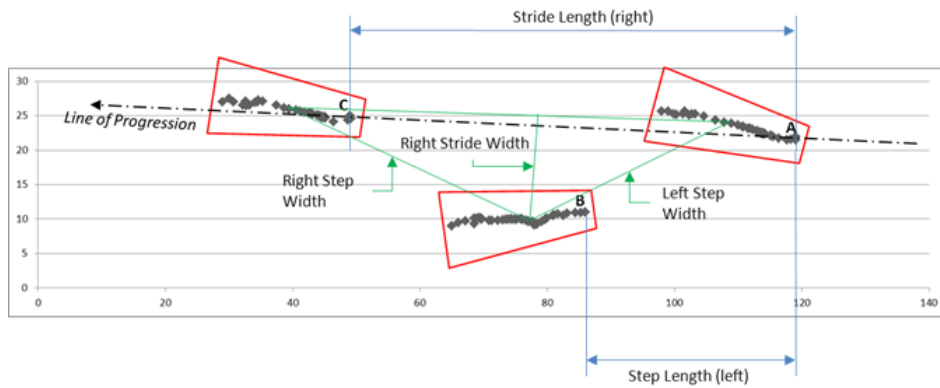


Fig. 2 Parameters identification on footfall

Based on these generated parameters during the whole walking cycle, it is possible to compute the standard deviation and coefficient of variations for stride time, stride length, stance time and base of support in order to get complementary information about the gait

(one cycle = stride). In addition, it is possible to record the footprints and respective pressure values stamped and also assess the respective CoP values. As presented by [11], the risk of falls can be measured by correlation with (i) “stride-to-stride” variability in velocity, (ii) variability in stride width and (iii) double-support time. This information is implicit in the CoP distribution of each foot. Based on the spatial and temporal parameters, the doctor is able to evaluate the gait and provide the diagnosis related to a specific problem.

Some other additional information can be extracted from the CoP's trajectory. Fig 2 shows three bounded regions representing the foot position extracted from the walking along the indicated line of progression, and the inner points are the CoPs trajectory for each one. The positions labeled as 'A', 'B' and 'C' are those of the heel of foot in its first point of contact. This is the pattern assumed by the Gaitrite system as reference to stride length and step length calculation [7]. The stride width and step width might be measured by Gaitrite and this relationship is also based on the heel position. Toe-toe measures instead of heel-heel measures have been chosen in specific cases for calculating the gait parameters as in [16], because the measure of lengths of step and stride depends on how much pressure information exists in a specific region of the footprint. In our research we are adopting a centroid reference as indicated in Fig. 2 based on the method of [14].

The CoP values are exported by the Gaitrite system, and it is the point of location of the vertical ground reaction force vector in each foot [14]. In motion, a set of the CoP defines the trajectory of each foot while in the ground-contact during the time, based on its weighted average pressure values in the contact area. The CoP position values change in various directions under the foot according to the individual sway. Note that a different interpretation is given if the analysis is based on a sensor plate [17, 18] where the CoP is a calculated median value between feet ($CoP = CoG$). The assumption is different in our study because the CoPs are individual set data from each individual footprint ($CoP \neq CoG$). A graphical plot of its (x, y) coordinates gives a visual sense of foot movement during the footfall. To sum up, the CoP is the median representation of all the affected sensors along the footfall during the foot sway and it gives a foot signature map that describes the footfall behavior [18, 21].

2.2. Data normalization and repository

The Gaitrite system exports a file containing the individual trajectories of the pressure centres of each foot of a given patient. Each patient is submitted to five standardized testing protocols which are: 1) normal walking; 2) slow walking; 3) fast walking; 4) normal walking in conjunction with cognitive task (cognitive task is the patient speaking aloud the countdown by 2 in 2, starting with 50 or 30 depending on the condition of the patient); and 5) normal walking associated with motor activity (motor activity is the patient carrying a glass cup with 2/3 filled with water). Each test combines two runs, i.e. going and coming back on the walkway.

The generated file contains the gait data tabulated in ASCII format as in Fig 3. The file can be imported as an Excel spreadsheet whose columns are: the parameter 'obj', i.e.. the sequential number of each foot; the 'time' column records the time (in ms) in which each sensor has been activated; column 'L/R' is the respective foot in ground contact, where '0' represents the right foot, and value '1' to the left foot; columns 'X' and 'Y' contain the coordinates of each centre of pressure (COP) forming part of the sensor activation trajectory line during the contact of each individual foot on mattress.

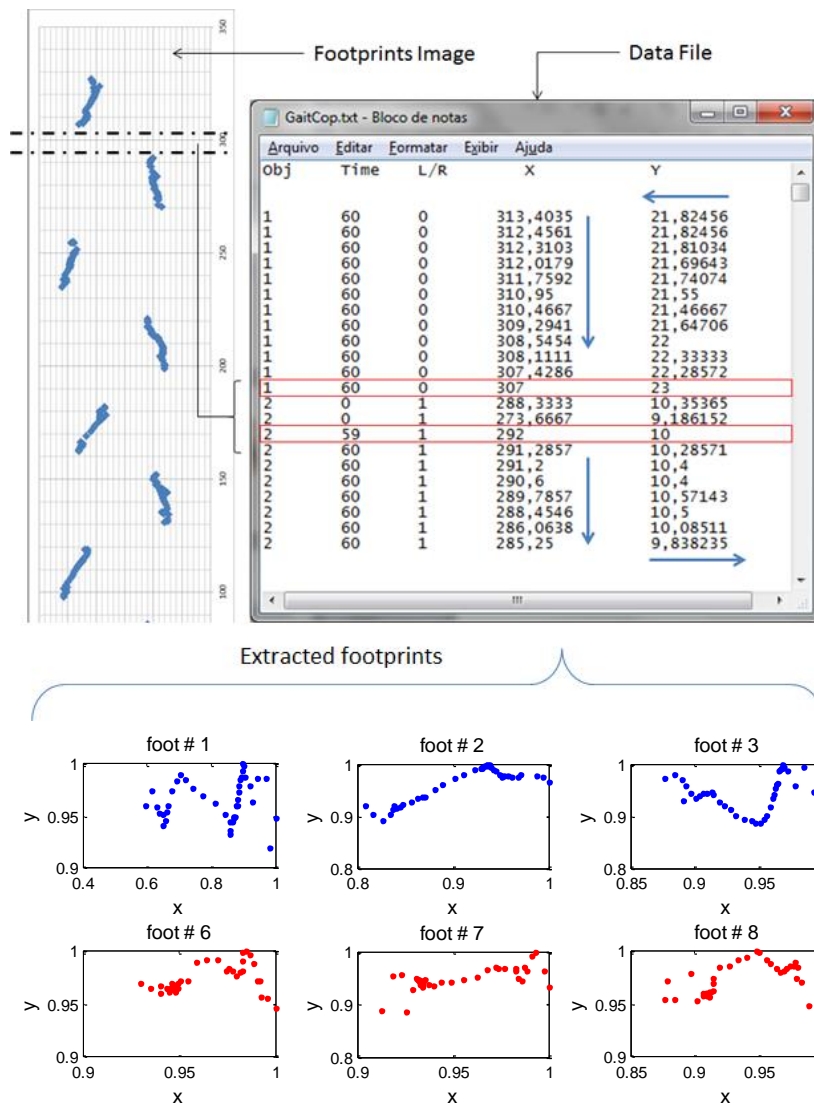


Fig. 3 File model to data exchange from GaitRite® and respective extracted footprints

From the CoP file we obtain all separate footprints graphs [23], as indicated by blue (right foot) and red (left foot) plots in Fig. 3. Each footprint represents the respective foot signature compound by their X and Y coordinates normalized to [0-1] range. A repository database is organized by all the spreadsheets containing walking information for all patients.

2.3. Data mining procedure

As described by Mirkovic in [25], data mining is a practical procedure based on clustering and it is widely applied to group selection. It represents roles to find patterns and also implements methods in order to describe the structural patterns in data repository. In our context, we apply this functionality in order to classify healthy or non-healthy individuals by their CoP footprints. The data is taken from a sample set with individual trajectories of the CoPs of each foot for a given patient. We expect that the output should give us a prediction about new samples. In our approach we have applied the RapidMiner® Studio software [24] to the data mining because it is an open source tool; it also has applicability for a wide range of problems. The steps of implementation are shown in Fig 4.

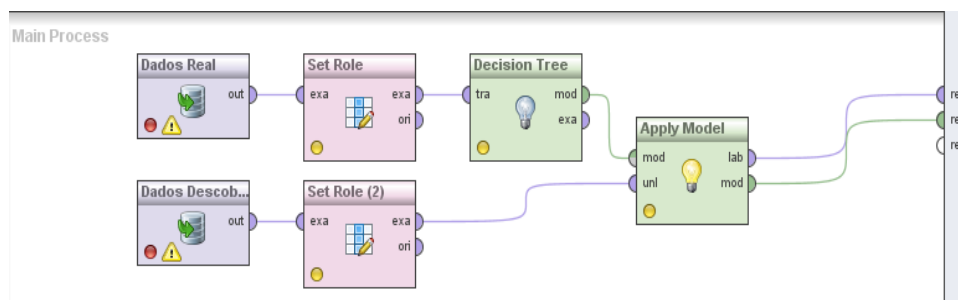


Fig. 4 Steps implementation of mining process

As depicted in Fig. 4, the mining process implementation has the following operations as defined in [24]:

2.3.1. Retrieves

Each ‘Retrieve’ activity is an operator with reading function of an object from the data repository. Thus, we can get data / metadata like subject and his particular set of parameters. For our process of the gait analysis we have two Retrieves where one ‘retrieve’ is the known data and the other is the testing data. The training set contains X and Y coordinates of the footprints of all subjects, as plotted in Fig. 5.

2.3.2. Set roles

A ‘Role’ operator is used to change the role for each of the attributes. The role of an attribute reflects its categorization and embeds some specific task. The special roles are: label, id, prediction, cluster, weight, and batch. For our proposal, we have two Roles. A first operator Set Role is to assign the role identifier to the attribute "patient" just to its own identification as well as the “healthy” attribute is assigned as a label for it. The second Role Set is related to testing data and it carries the attribute "Patient" only. (Note that the healthy or non-healthy is a further output).

2.3.3. *Decision tree*

It is a graphical representation tool of the searching model composed by nodes and leafs [22]. The idea behind the decision tree is to create the classification model whose function is to predict the classification of a value. Each inner node corresponds to one of the input attributes and each leaf node represents a value of the label attribute (the path from the root to the leaf). Thus the decision tree implements the model.

2.3.4. *Apply model*

The model is first trained through a known sample, and the related information is learnt by the model. After this, the model (decision tree) can be applied to another set of values for prediction. This operation generates the processed output.

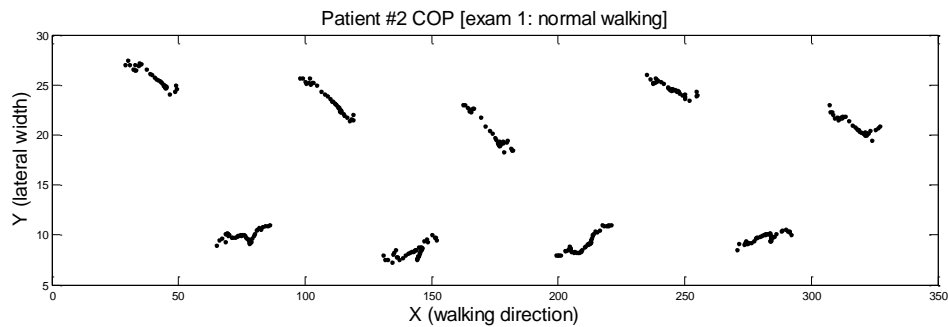


Fig. 5 An example of footprints along walkway (from data repository)

3. RESULTS AND DISCUSSION

A testing condition is implemented in order to evaluate the proposed classification mechanism. The range of test includes the total of 22 healthy and 7 non-healthy patients through the amount of almost 200 footprints. Fig. 6 shows an example of the CoP distribution under one individual foot for both (a) healthy and (b) non-healthy cases.

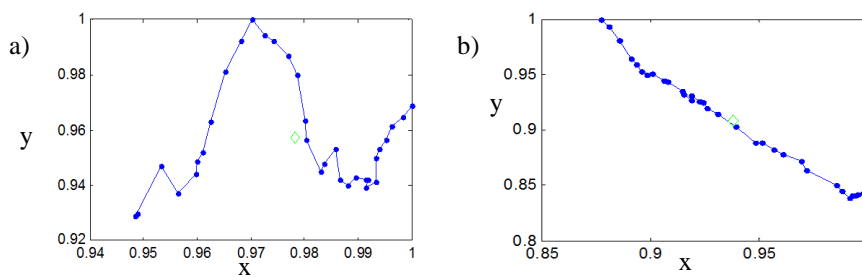


Fig. 6 Example of the CoP distribution of one foot, for (a) healthy and (b) non-healthy individual

In Fig. 6, the points are, respectively, X and Y Cartesian coordinates of the COP trajectory. The normal walking protocol as experienced and the respective decision tree generated by the data mining is presented in Fig. 7.

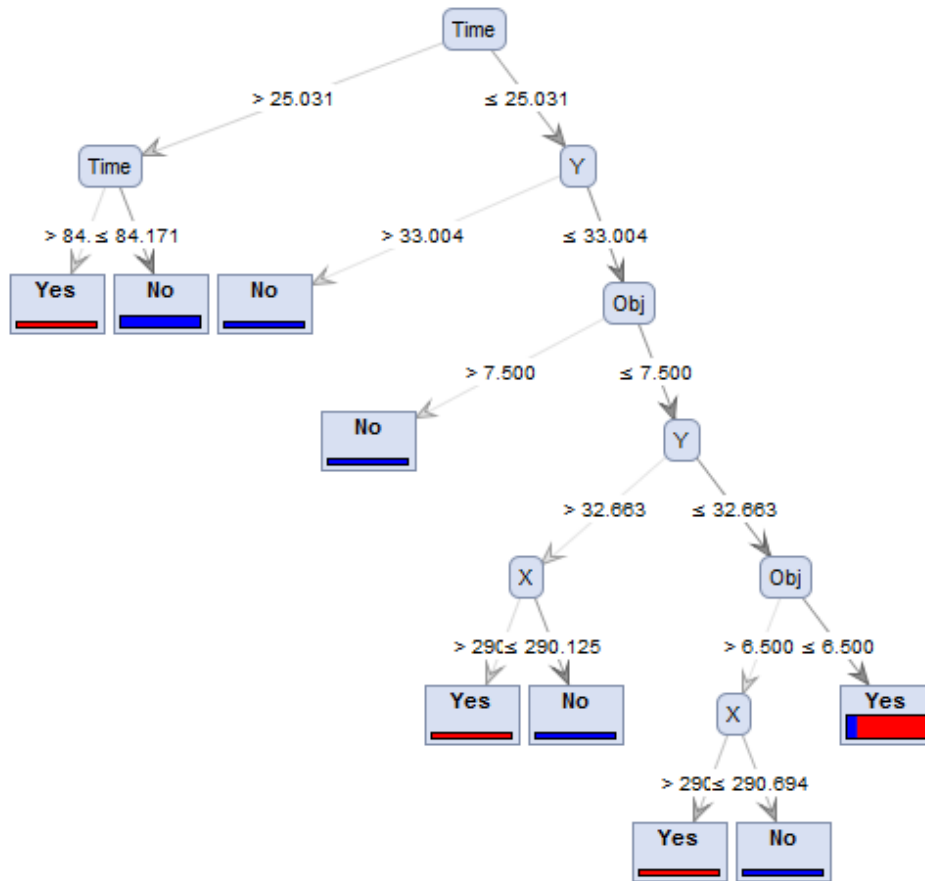


Fig.7 Decision tree to pattern classification from footprint parameters

The parameters of the searching model are, the contacting 'time' of each foot (heel-toe) on ground and object 'obj' (right or left foot) as the nodes in the decision tree. The leaves are 'X' and 'Y' coordinates as a deviation in the CoP trajectory in each foot on walkway (as normalized data examples previously presented in Fig. 6). Fig. 8, in (a) presents the comparative graphical representation about the gait pattern classification executed between the real data and the respective mining classification.

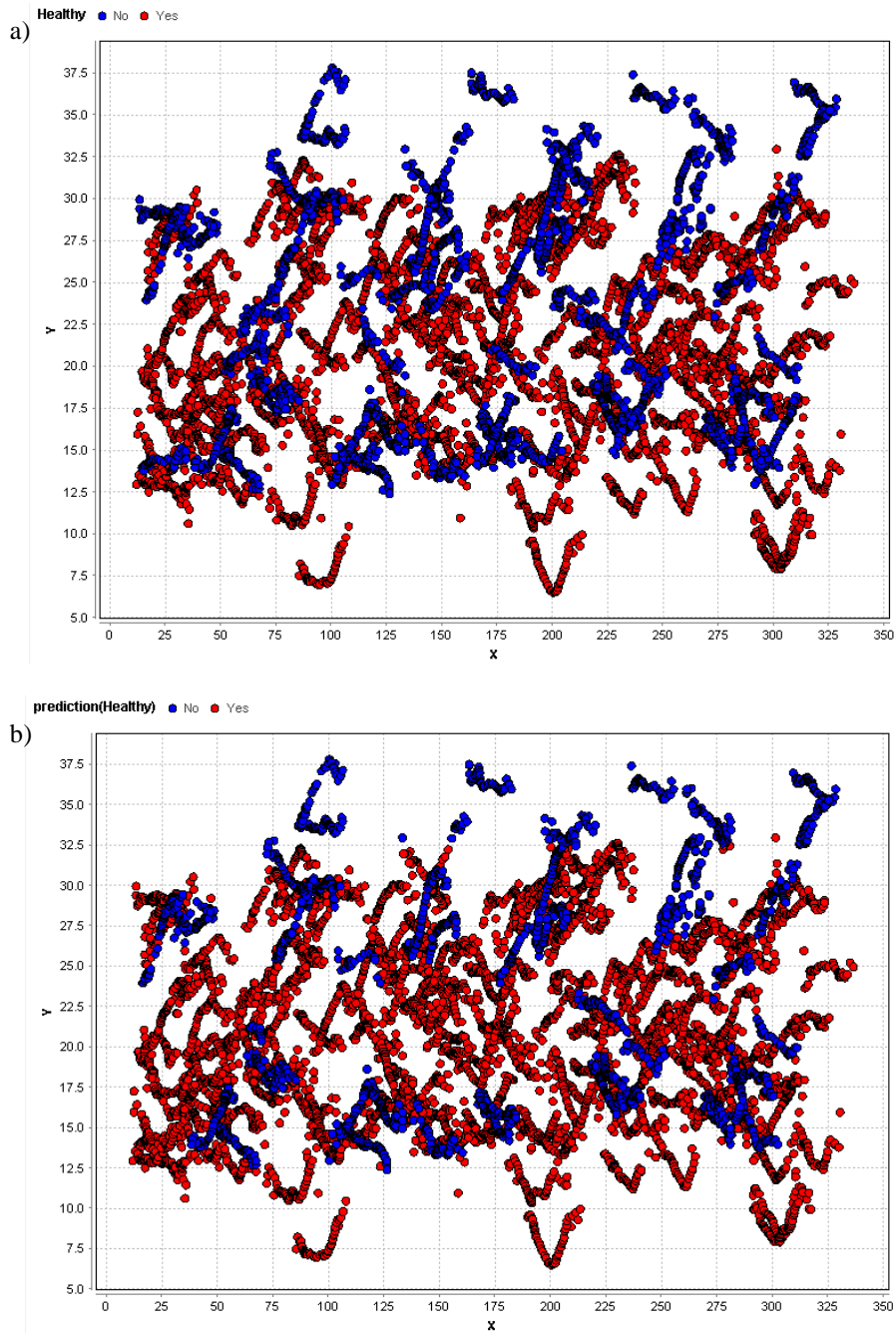


Fig. 8 (a) Real data from all patients; (b) Classification for testing subject

From graphs in Fig. 8, in (a) presented are in red all the footprint plots for the healthy people set and the blue are footprints for a non-healthy group. The red or blue points are the trajectory of each patient along the walkway. The representation is based on a pair of coordinates of their respective CoP trajectories. The graph in (b) shows the footprints classification result based on the decision tree. In this case all the parameters, i.e. 'time', 'object' and 'X', 'Y' CoP coordinates for each patient are classified according to the decision tree. In the same way the graph shows the footprints in red and blue for healthy and non-healthy subjects, respectively.

Thus it is possible to have a visual interpretation of the pattern classification. In graph (b) it possible to see that the resulting data for a testing subject are very similar to the classification ones. Surely, we can generally understand how the classification works; however, the visual sense of the pattern is the only information possible to get from the graphs. The displacement measurement cannot be estimated and an additional mechanism should be implemented to improve this evaluation.

3. CONCLUSIONS

The text presents a proposal of the method developed to provide the classification of individuals in two possible conditions concerning their gait, so called healthy and non-healthy subjects, through analysis of their gait patterns. The centre of pressure (CoP) trajectory in each foot is calculated by the GaitRite system and the parameters (time, footprint position) can be used to define gait characteristics. A data mining approach is implemented as a computational tool for the support in the gait pattern classification. The main application of method is to assist the clinical diagnosis and respective treatment evaluation. A decision tree is modeled in order to provide the classification of objects according to the known training group. The example has shown that a selected patient is classified according its condition, i.e., his gait behavior is mapped inside the correct group. Then, the tested features have contributed to the classification of individuals regarding their similarity, and as demonstrated, the proposal is a promising tool in the gait analysis context. For future works, we want to improve the classification (decision tree) for patient classification by their associate deceases adding machine learning strategies. Different kinds of deceases have an influence on the gait pattern (due stroke, surgery, etc.) and we still need to process more specific patterns by a reasoning system in order to assist gait treatments.

Acknowledgements: *The authors would like to thank the Zitha Senior and LIST for clinical data providing. Also the Pontifical Catholic University of Parana (PUCPR) and the CAPES Brazilian Grant program BEX 9584/11-0 by financial support.*

REFERENCES

1. Panzer, V.P., Wakefield, D.B., Hall C.B., Wolfson, L.I., 2011, *Mobility Assessment: Sensitivity and Specificity of Measurement Sets in Older Adults*, Arch Phys Med Rehabil., 92, pp. 905-912.
2. Robinovitch, S.N., Feldman, F., Yang, Y., Schonnop, R., Leung, P.M., Sarraf, T., Sims-Gould, J., Loughin, M., 2013, *Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study*, Lancet, 381, pp. 47-54.

3. Minetti, A. E., Cisotti, C., Mian, O.S., 2011, *The mathematical description of the body centre of mass 3D path in human and animal locomotion*, Journal of Biomechanics, 44, pp.1471–1477.
4. Sheehan, K.J., Greene, B.R., Cunningham, C., Crosby, L., Kenny, R. A., 2014, *Early identification of declining balance in higher functioning older adults, an inertial sensor based method*. Gait & Posture, 39, pp.1034–1039.
5. Sparto, P.J., Jennings, J.R., Furman, J.M., Redfern, M.S., 2014, *Lateral step initiation behavior in older adults*, Gait & Posture, 39, pp. 443–448.
6. Yeh, H.C., Chen, L.F., Hsu, W.C., Lu, T.W., Hsieh, L.F., Chen, H.L., 2014, *Immediate Efficacy of Laterally Wedged Insoles With Arch Support on Walking in Persons With Bilateral Medial Knee Osteoarthritis*, Archives of Physical Medicine and Rehabilitation, 95(12), pp. 2420–2427.
7. GAITRite Electronic Walkway, 2015, Technical Reference (WI-02-15), pp. 31-50.
8. Tsai, Y.C., Hsieh, L.F., Yang, S., 2014, *Age-related changes in posture response under a continuous and unexpected perturbation*, Journal of Biomechanics, 47, pp. 482–490.
9. Taylor, M., McEwen, D., Goubran, R., Finestone, H., Knoefel, F., Sveistrup, H., Bilodeau, M., 2012, *Assessing Standing Stability of Older Adults using Pressure Sensitive Arrays*, the IEEE International Symposium on Medical Measurements and Applications, pp. 1 – 5.
10. Taylor, M., Goubran, R., Knoefel, H., 2012, *Patient Standing Stability Measurements using Pressure Sensitive Floor Sensors*, the IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1275 – 1279.
11. Milankovic, I., Rankovic, V., Peulic, M., Filipovic, N., Peulic, A., 2015, *Diagnosis of Lumbar Disc Herniation using Multilayer Perceptron Neural Network*, the 5th International Conference on Information Society and Technology, ICIST 2015, pp. 210 – 213.
12. Bih-Jen, H., Fong-Chin, S., 2014, *Effects of Age and Gender on Dynamic Stability During Stair Descent*, Archives of Physical Medicine and Rehabilitation, 95, pp.1860–1869.
13. Saito, I., Okada, K., Nishi, T., Wakasa, M., Saito, A., Sugawara, K., Takahashi, Y., Kinoshita, K., 2013, *Foot Pressure Pattern and its Correlation With Knee Range of Motion Limitations for Individuals With Medial Knee Osteoarthritis*, Archives of Physical Medicine and Rehabilitation, 94, pp. 2502–2508.
14. Chockalingam, N., Bandi, S., Rahmatalla, A., Dangerfield, P.H., Ahmed, E.N., 2008, *Assessment of the centre of pressure pattern and moments about S2 in scoliotic subjects during normal walking*, Scoliosis Journal, 3(10), pp.1 – 6. (doi: 10.1186/1748-7161-3-10)
15. Hernández, A., Silderb, A., Heiderscheit, B.C., Thelen, D.G., 2009, *Effect of age on center of mass motion during human walking*, Gait & Posture, 30(2), pp.217–222.
16. Sorsdahl, A.B., Moe-Nilssen, R., Strand, L.I., 2008, *Test–retest reliability of spatial and temporal gait parameters in children with cerebral palsy as measured by an electronic walkway*, Gait & Posture, 27, pp. 43–50.
17. Winter, D.A., 1995, *Human balance and posture control during standing and walking*, Gait & Posture, 3, pp.193-214.
18. Tseng, I.F., Chern, J.S., 2008, *Bilateral Foot Center of Pressure during Trunk Forward Bending and Reaching*, the IEEE International Conf. on BioMedical Engineering and Informatics, pp. 566 – 571.
19. Tung, J.Y., Gage, W.H., Zabjek, K.F., Maki, B.E., McIlroy, W.E., 2011, *Frontal plane standing balance with an ambulation aid: Upper limb biomechanics*, Journal of Biomechanics, 44, pp.1466–1470.
20. Caderby, T., Yiou, E., Peyrot, N., Begon, M., Dalleau, G., 2014, *Influence of gait speed on the control of mediolateral dynamic stability during gait initiation*. Journal of Biomechanics, 47, pp. 417–423.
21. Balasubramanian, C., Gouelle, A., 2015, *The Gait Variability Index, a new composite measure of gait variability, decreases with aging*, The 25th Annual Meeting of the Society for the Neuro Control of Movement. Charleston, South Carolina, USA.
22. Larose, D. T., 2014, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, .
23. Manassah, J. T., 2013, *Elementary mathematical and computational tools for electrical and computer engineers using MATLAB*. CRC Press, 2013.
24. RapidMiner 6 - User Manual. Dortmund, Germany (2014). Available in: www.rapid-i.com (last access 10.10.2015.)
25. Mirkovic, D., Lukovic, I., Obrenovic, N., Rogic, D., 2015, *A Framework for Comparative Analysis of Data Mining Algorithms*, the 5th International Conference on Information Society and Technology, ICIST 2015, pp. 49-54.