A Proposal for Footprints Classification in Gait Patterns based on the trajectory of Centres of Pressure

Marcelo Rudek*, Nicoli Mariá Silva *

* Pontifical Catholic University of Parana, PUCPR, Production and System Engineering Graduate Program – PPGEPS, Curitiba, Brazil

marcelo.rudek@pucpr.br, nick_nicoli@hotmail.com

***Abstract*—The pressure map measured under foot by the contact with the ground is an essential tool to provide information how everyone is moving. Among all spatial and temporal data whose values can be measured by pressure gathering systems, the trajectory of the Centre of Pressure (CoP) in each footprint permit us to see a kind of foot signature from an individual. Thus, it is possible to get some level of patterns' grouping just based on similarity among all CoP trajectories. The problem is the large amount of data for comparing among all pressure maps that become unfeasible a visual inspection task by humans. From the engineering point of view, this research proposes a method applied to gait pattern classification based on Knowledge Discovery in Database (KDD), where the objective is to understand if exists specific patterns which permits to group and to classify footprints by an automatic way. The text also presents an example of application of the method. The database contain data from two testing groups, compound by healthy and unhealthy younger adults and elderly people. The health condition of each one was previously known, and after application of method, the decision tree classified the footprints groups with precision of 95,3% of confidence. The presented method demonstrated that is feasible to define a computational classification criterion based on CoP trajectory in order to assist gait analysis.**

I. Introduction

The movement generated by walking is extremely necessary for humans' autonomy in its daily activities at any stage of their life. Over time, problems associated with walking increase the difficulty in locomotion and balance disorder. The engineering field provides measuring equipment and software tools which allow to collect data of how a person is moving. For instance, such tools enhance the analysis procedures to gait evaluation, serving as decision maker in order to support the clinicians. Thus, it makes possible to transform qualitative indicators, which information are difficult to compare and measure, in quantitative ones, which data can be handled by the computational expert methods.

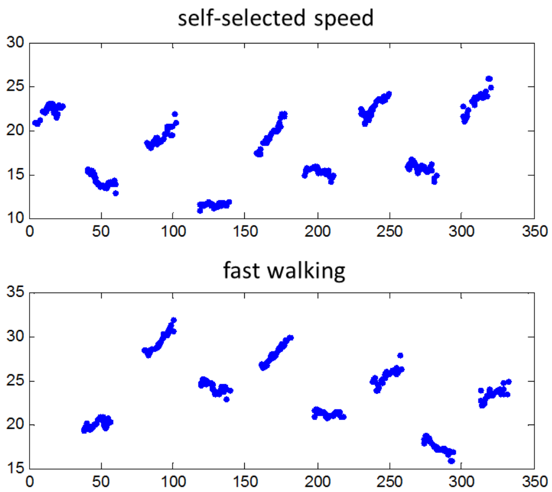
Monitoring technologies based on biometric signals have improved this area, mainly contributing with rehabilitation and to early detection of diseases [1-6]. Tools such cameras, force plates and insole electronic devices are some of the most recent technologies applied to gait analysis research [7,8]. In addition, the walkway mat provides pressure maps generated from contact of the feet on device during walking [9] whose pressure levels are displayed as images from respective areas in the foot with major or minor pressure.

The application of these devices in daily practice of clinical studies generate a large amount of data, and even though a clinician has large experience, it can spend a lot of time and resources to analysis and diagnoses. As well as, the history of data acquisitions can be kept in a database for further studies in treatment's progression of someone, or also to provide comparative analysis among different pathologies and their respective treatments.

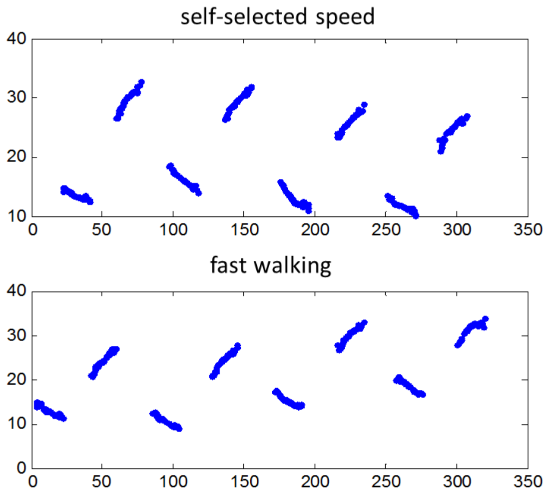
A. The statement of problem

According to [10,11], the Center of Pressure (CoP) values distributed on the footprint's image provides an estimative of the position of the center of gravity of an individual projected on the ground, and it changes according the individual is moving or swinging. Thus, as stated by [12] the quantitative spatiotemporal measure of CoP has been repeatedly proven as a robust indicator of quality in functional stability.

The GaitRite® System is a specialized tool for indoor gait analysis and it provides compiled results based on spatial and temporal data with large confidence. Also, it shows the footprints' image sequence with their respective pressure levels. However, additional information from CoP's influence to gait evaluation, as for instance the respective trajectory coordinates in each foot, must be obtained by an exported file. The sequence of the CoP trajectory in each foot can be rebuilt and displayed as a graph. The Fig.1 shows two samples of data collected from two individuals under standardized exams conditions called "protocols". By an intuitive visual observation in Fig.1, for both individuals (a) and (b) it is possible to see different patterns in the alignment and grouping of CoP footprints between them. However, the pattern looks like remains similar between distinct velocities to the same individual. Since that it is possible to observe a kind of "walking signature" from CoP map, it can be inferred that the CoP may be indicative of a specific behavior pattern as a people walks. Certainly, due to the large number of footprints that can be obtained from each examination for several patients, the inspection by humans is practically unfeasible. Thus, the main question is how to define the parameters whose values can be applied to create an automated classifier to classify individuals according to the gait patterns similarities.



(a)



(b)

Figure 1. Two samples of the CoP footprints from two different adults.

A Knowledge Discovery in Database (KDD) process through data mining strategy can be applied to allow an autonomous classification for different groups of data, i.e. gait patterns. In this way, we apply a classifier based on the supervised learning technique represented by a decision tree. A training data set was previously explored in [13] and [14] with a classifier to all the CoP coordinates for all exams. In fact, considering that each exported CoP file contains, on average 7 footprints, and each one containing about 50 Cartesian coordinate pairs (X, Y), then the total volume of data for tests were approximately 122,500 coordinates for one person. In that case, the result obtained in [13] was not conclusive due a lack of a pre-processing level in KDD.

Now, in this research, instead to consider all Cartesian coordinates, we evaluate only some specific distances, as the distances between the heels center and metatarsus center from each step. Then, a pre-processing stage to proceed this calculation was included to improve the quality of data, and thus it was possible to reduce the amount of data to only about 1,630 instances for all exams. In this case, we are verifying if the distances variation among the CoP footprints are enough to proceed a correct classification of patterns.

B. Measurement Device

The gait consists in the body movement by alternating the support of its weight between the two legs, characterized as rhythmic movements to the progressive locomotion ahead [15-17]. A walkway device can be applied to register this movement and the Knowledge Discovery in Database (KDD) offer the computation support for pattern classification. The Gaitrite is an indoor walkway system, whose apparatus consists of a rubberized mat with 18,432 pressure sensors built into an active area of 4.3 meters of length by 0.6 meters in width, as presented in Fig.2. The measurement system [9] and [18-20] defines spatial and temporal parameters used for gait analysis (for example step and stride length/width and times, toe in/toe out position, gait cycle, velocity and cadence among others).

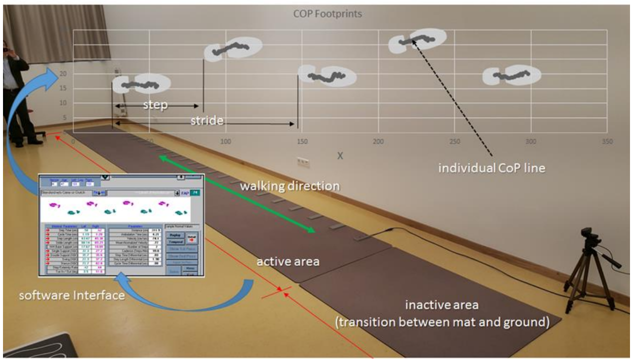


Figure 2. Walkway apparatus and its main parameters.

According the patient walks over the active area, the mechanical pressure of the feet on the carpet activates the sensors and the system records the geometry and relative configuration of each footprint.

C. Knowledge Discovering in Database (KDD)

Knowledge Discovery in Database (KDD) according to [21] is an interactive and iterative process with several steps that require decisions to be made by the user. Reference [6] shows that the process of knowledge discovery in databases can be divided into three basic operational steps: Pre-processing, Data Mining and Post-processing. The first step covers the preliminary steps of collecting or accessing the database to the organization and processing of the data. The second one is Data Mining that aims to discover patterns, which can be represented in different models from machine learning algorithms [22]. Subsequently, the third stage deals with post-processing and refers to evaluative activities, with the intention of analyzing the quality of the patterns in relation to the initial problem.

II. Proposed Method

The proposed method seeks to chain the necessary steps to include the CoP information in the gait analysis, complementing the analysis generated by the GaitRite system. Thus, in addition to the individualized analysis based on spatial and temporal data for each patient, which is a characteristic of the system, the method allows inter-individual comparison to be performed by analyzing the CoPs image patterns of all individuals observed together. According Fig.3, the proposed method has three main layers, such as Data Gathering & Analysis, Data Processing and Knowledge Discovering in Database (KDD).

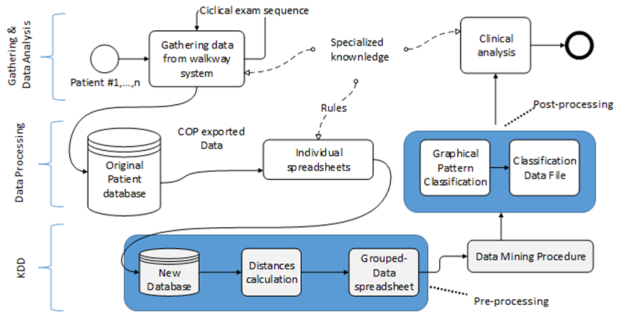


Figure 3. Proposed method improved from [13].

The initial step is the data gathering performed by a physician within a sequence of the clinical exams according five specific protocols, defined in [6,13] as:

- •protocol (1) - normal walking, with self-selected patient speed;
- •protocol (2) - slow walking, given the instruction to the patient / volunteer walk at the lowest speed without stopping;
- •protocol (3) - fast walking, orientation to walk as fast as possible, but without running;
- •protocol (4) - cognitive dual task walking is performed by counting (verbalizing) in decreasing order by two;
- •protocol (5) - normal gait associated with motor activity, carrying a glass cup filled with 2/3 of water.

Thus, all temporal and spatial data collect by [9] from each one individual is stored in the system database in the second level. The both levels interact to store and provide data, as for example, to export CoP files. A specialist based on own experience and knowledge apply its rules for interpretation of information. Due the large amount of data, the proposed method added the KDD level in the methodology as a support to data analysis, whose process embedded two main tasks, respectively pre-processing and data mining.

A. Pre-processing

According to [6], the pre-processing is the needed step for the data preparation to Data Mining and includes abstraction, organization and treatment of the data. In our context, the pre-processing task is applied to reduce the amount of data. Thus, in this phase it investigated if the distance values among CoPs footprints can be a parameter for classification, instead use a wide amount of data from all CoP cartesian coordinates.

The new calculated parameters are presented in Fig.4. For each footprint, there are two distinct points, which are respectively the heel center (HC) and the center of metatarsus (CM). These points were used to calculate the distance in the step by two ways: (i) distance between the HC to the center of the heel of the next foot (HC_HC); and (ii) distance from to the heel center of the current foot to center of metatarsus of the next foot (CM_HC). We notice that these distances are not the same (step and stride width) already provided by [9].

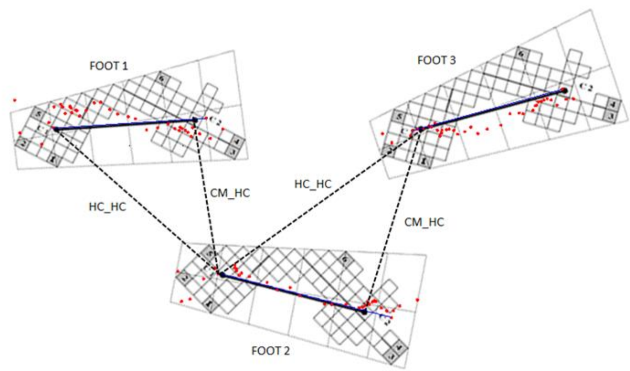


Figure 4. Definition of distance parameters nomenclature.

B. Data Mining

According [22-24] the data mining technique by Supervised Learning is the one that discovers the response variable based on the forecast variables. The Weka software [25], implements the data mining by supervised learning technic and it applies a classification algorithm for classifying records with previously defined categorical labels (classes) called J48 classifier. For the data mining, it is applied a “training data set” with well-known data categorization, as for example young, elderly, healthy, unhealthy and so on. From this data set, Weka builds a decision tree based on distances HC and CM. In order to evaluate the classifier performance, the decision tree is applied to a different “test dataset” with unknown categorization. A so called “confusion matrix” [6] shows the prediction and false positives to that test. Thus, data mining predicts the correct class of individuals conditions (for instance, healthy or unhealthy).

III. Example and Discussion

An example to application of the method considers two different datasets, as:

Training Data Set -this group contains 29 individuals, where it has 22 volunteers, young adults considered healthy (without pathology known) and the remain 7 are elderly patients submitted to total hip arthroplasty or to endoprosthesis hip surgery. All CoP data from those five different protocols are distributed in 282 files.

Test Data Set - This data set contains measurements taken from 68 individuals, all considered healthy with distinct ages among young and elderly adults. In this case the data were collected for only the protocol (1) -"self-selected speed" speed by the patient himself, generating just one CoPs exported file per individual.

In an intuitive way, one can expect that the patterns of the two groups may have different characteristics, since they deal with very different situations, because possibly the pattern of the unhealthy elderly and healthy adult pattern should be easily identified. By this example, it is expected that the classification after decision tree (quantitative analysis) corresponds to our intuitive analysis (qualitative analysis). Another question involved is also to verify the correlation between the different protocols. Thus, in this example it is sought to verify that the gait in different speeds (protocols 2 to 5) maintains the same pattern of the protocol 1 (self-selected speed) as criterion.

The decision tree produced the confusion matrix presented in Table I. The quality evaluation of decision tree was measured by the parameters S=Sensitivity (Recall), P=Precision and F = *F* score (harmonic mean) as defined in [6] and [26-28].

Among 1027 instances of healthy people, the model made 997 correct predictions. Also, among 497 instances of unhealthy people, 456 were classified correctly. In total 71 instances were incorrectly classified compared to the original file to both groups, thus the overall accuracy rate is 95.34% according harmonic mean F-measure [26]. In order to evaluate the accuracy of the discovered classifier, a confusion matrix was constructed from the Test Data Set, as a different database from that used to discover the classifier.

Table I. Confusion Matrix for the decision tree applied to the Training Data Set.

Training Data Set		Predicted Class		Total of Instances
		Healthy	Unhealthy	
Actual Class	Healthy	997	30	1027
	Unhealthy	41	456	497

S	P	F	Accuracy Rate (%)
0.971	0.961	0.966	95.3

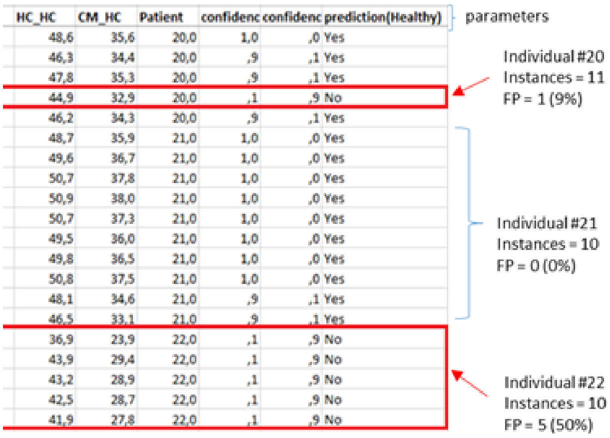
The result is presented in Table II. Considering that all individuals were healthy-classified, among 605 instances, 69 instances are false-positives, that is, 88.6% of accuracy rate in the classification. It signifies that 69 calculated distances are compatible with an unhealthy condition.

Table II. Result of Confusion Matrix applied to the Test Data Set

Test Data Set		Predicted Class		Total of Instances
		Healthy	Unhealthy	
Actual Class	Healthy	536	69	605
	Unhealthy	-	-	-

S	P	F	Accuracy Rate (%)
0.886	1.000	0.940	88.6

As presented in Fig.5, the individual (patient) #20 is healthy in its known condition, but it has one instance predicted as unhealthy. For sure, just one instance identified as false positive is not enough to qualify his condition as unhealthy. In addition, the individual #22 has five instances predicted as unhealthy, but his known condition was healthy. Then maybe this individual should be examined to observe its real condition. Thus, the number of instances incorrectly predicted from the initial condition, can define a new parameter to make decision about the patient condition. The improvement to classification in terms of individuals (Patients/Volunteers) is through a new parameter called threshold. It defines a limit value dependent of the number of false-positive instances for each individual. The Table III presents the relationship between the training and test data sets for two specific levels for the threshold, and their respective accuracy by the number of false-positives.



HC_HC	CM_HC	Patient	confidenc	confidenc	prediction(Healthy)
48,6	35,6	20,0	1,0	,0	Yes
46,3	34,4	20,0	,9	,1	Yes
47,8	35,3	20,0	,9	,1	Yes
44,9	32,9	20,0	,1	,9	No
46,2	34,3	20,0	,9	,1	Yes
48,7	35,9	21,0	1,0	,0	Yes
49,6	36,7	21,0	1,0	,0	Yes
50,7	37,8	21,0	1,0	,0	Yes
50,9	38,0	21,0	1,0	,0	Yes
50,7	37,3	21,0	1,0	,0	Yes
49,5	36,0	21,0	1,0	,0	Yes
49,8	36,5	21,0	1,0	,0	Yes
50,8	37,5	21,0	1,0	,0	Yes
48,1	34,6	21,0	,9	,1	Yes
46,5	33,1	21,0	,9	,1	Yes
36,9	23,9	22,0	,1	,9	No
43,9	29,4	22,0	,1	,9	No
43,2	28,9	22,0	,1	,9	No
42,5	28,7	22,0	,1	,9	No
41,9	27,8	22,0	,1	,9	No

Figure 5. Resulting spreadsheet after decision tree.

Table III. Training and testing for different threshold.

Group	Number of patients	Threshold	FP	Accuracy (%)
Training	29	0.0	7	75,86
Test	68	0.0	9	86,76
Training	29	0.5	0	100,0
Test	68	0.5	6	91,18

Considering individuals (patients / volunteer) instead the instances (distances), the classification rate reaches 100% assertiveness to the training data set and grows up from 88.6% to 91.1% to the testing data set for the threshold equal 0.5.

According to [29-30], people who have submitted to hip surgery have an asymmetrical gait pattern and it have a shorter stride length compared to people not submitted to this procedure. The method is able to show the incorrectly classified person (false positives) presents short distances HC_HC and CM_HC, being compatible to a person with a pathological condition.

IV. Conclusion

A method for the pre-processing in KDD strategy was proposed based on distances among selected points in footprints. Those all distance values were stored in a training dataset with its respective known about the individual condition (healthy or unhealthy). After, a decision tree was applied inside initial dataset, and it was obtained an accuracy rate of 95.3% for those data. It significates that the decision tree was able to classify each instance correctly in ~95% of cases in your own dataset. In this test, we can infer that the patterns from different walking velocities are similar.

In order to validate the decision tree, the same model was applied to another testing dataset. In this sense, the model was calibrated considering all protocols from the Training Data Set. After, it was applied to the dataset which contains just the protocol (1) – self-selected walking. There were 605 instances (distances calculated from 68 patients/volunteers), whose false-positives were 69 instances incorrectly classified. These false-positives distances belong to nine distinct individuals. Those individuals are people with an average age about 75 years old. Also, only three instances were from one non-elderly person. In conclusion, the decision tree based on distances (HC and CM) found the elderly people among all set of data.

The interpretation for this data is that it is possible to separate the elderly and young groups independent of healthy or unhealthy condition. This study shows a possible path to future researches to explore the spatial parameters from CoP to distinguish different diseases in similar groups.

Acknowledgment

The authors would like to thank the Luxembourg Institute of Science and Technology (LIST) and Department of Research and Development, Zitha Senior - Centre on Memory and Mobility Michel Rodange, Luxembourg, by technical support and data collecting and to Pontifical Catholic University of Paraná by financial support.

References

[1] M. T. Martinez and P. De Leon. Falls Risk Classification of Older Adults Using Deep Neural Networks and Transfer Learning. IEEE Journal of Biomedical and Health Informatics,2019.

[2] O. Dehzangi, B. A. Bache and O. Iftikhar. Activity Detection using Fusion of Multi-Pressure Sensors in Insoles. 24th International Conference on Pattern Recognition (ICPR), pp. 3315 – 3321, 2018.

[3] S. N. Robinovitch et al. Video capture of the circumstances of falls in elderly people residing in long-term care: an observational study. The Lancet, v. 381, n. 9860, p. 47-54, 2013.

[4] K. J. Sheehan, et al. Early identification of declining balance in higher functioning older adults, an inertial sensor based method. Gait & posture, vol. 39, n. 4, pp. 1034-1039, 2014.

[5] Ö. F. Ertuğrul et al. Detection of Parkinson's disease by Shifted One Dimensional Local Binary Patterns from gait. Expert Systems with Applications, vol. 56, pp. 156-163, 2016.

[6] N. M. Silva, Proposta de um método de classificação de padrão de marcha baseado na distribuição dos centros de pressão. Dissertação De Mestrado – PUC-PR, Curitiba - PR, 2017.

[7] C. Balasubramanian, and A. Gouelle. 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, 2015.

[8] H-C. Yeh, et al. Immediate Efficacy of Laterally Wedged Insoles With Arch Support on Walking in Persons With Bilateral Medial Knee Osteoarthritis. Archives of physical medicine and rehabilitation, v. 95, n. 12, pp. 2420-2427, 2014.

[9] GAITRite Electronic Walkway Technical Reference (WI-02-15) Rev. P. 2015.

[10] M. Taylor et al. Assessing standing stability of older adults using pressure sensitive arrays. Medical Measurements and Applications Proceedings (MeMeA), (2012) IEEE International Symposium on. IEEE, pp. 1-5.

[11] M. Taylor, R. Goubran and F. Knoefel. Patient standing stability measurements using pressure sensitive floor sensors. Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International. IEEE, (2012). pp. 1275-1279.

[12] I-F. Tseng, J-S. Chern. Bilateral Foot Center of Pressure during Trunk Forward Bending and Reaching. 2008 International Conference on BioMedical Engineering and Informatics. IEEE, (2008). pp. 566-571.

[13] M. Rudek et al. A data-mining based method for the gait pattern analysis. Facta Universitatis, Series: Mechanical Engineering, vol. 13, n. 3, pp. 205-215, 2015.

[14] N. M. Silva et al. "A Method for Automated Gait Pattern Classification." ISPE. Brazil, 2016.

[15] J. Rose et al. Muscle pathology and clinical measures of disability in children with cerebral palsy. Journal of orthopaedic research, vol. 12, n. 6, pp. 758-768, 1994.

[16] F. Prince et al. Gait in the elderly. Gait & Posture, vol. 5, n. 2, pp. 128-135, 1997.

[17] R. N. Kirkwood et al. Searching for stability as we age: the PCA-Biplot approach. International Journal of Statistics in Medical Research, vol. 2, n. 4, pp. 255, 2013.

[18] K. H. Cho, J. L. Hwang and H. L. Wan. Test–retest reliability of the GAITRite walkway system for the spatio-temporal gait parameters while dual-tasking in post-stroke patients. Disability and rehabilitation vol.37, n.6 (2015), pp. 512-516

[19] A. L. Mcdonough, et al. The validity and reliability of the GAITRite system's measurements: A preliminary evaluation. Archives of physical medicine and rehabilitation, vol. 82, n.3, pp. 419-425, 2001.

[20] B. Bilney, M. Morris and K, Webster. Concurrent related validity of the GAITRite® walkway system for quantification of the spatial and temporal parameters of gait. Gait & posture, vol. 17, n.1, pp. 68-74, 2003.

[21] U. Fayyad, G. Piatesky-Shapiro and P. Smyth. From data mining to knowledge discovery: an overview. Cambridge: AAAI Press; 1996.

[22] C. Tucker et al. A data mining methodology for predicting early stage Parkinson's disease using non-invasive, high-dimensional gait sensor data. IIE Transactions on Healthcare Systems Engineering vol.5. n4 (2015), pp. 238-254.

[23] W. Van Der Aalst. Process mining: discovery, conformance and enhancement of business processes. Springer Science & Business Media, 2011.

[24] J. R. Quinlan. C4.5 Programs for Machine Learning, San Diego, CA: Morgan Kaufmann Publishers. 1993.

[25] I. H. Witten and E. Frank. Weka. Machine Learning Algorithms in Java, pp. 265-320, 2000.

[26] Y. Ishikawa et al. Gait analysis of patients with knee osteoarthritis by using elevation angle: confirmation of the planar law and analysis of angular difference in the approximate plane. Advanced Robotics vol.31, iss.1-2,2017.

[27] D. M. W. Powers, Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation. School of Informatics and Engineering, Flinders University of South Australia, Technical Report SIE-07-001, 2007.

[28] T. Fawcett. An introduction to ROC analysis. Pattern Recognition Letters, vol.27, pp. 861–874, 2006.

[29] M. Petersen et al. Gait analysis after total hip replacement with hip resurfacing implant or Mallory-head Exeter prosthesis: a randomized controlled trial. Orthopaedic Proceedings, (2010), pp. 528-528.

[30] Z. He ; T. Liu ; J, Yi, A wearable sensing and training system: towards gait rehabilitation for elderly patients with knee osteoarthritis. IEEE Sensors Journal, pp.1, 2019.