2018 IEEE International Conference on Systems, Man, and Cybernetics

Gait Analysis Based Approach for Parkinson's Disease Modeling with Decision Tree Classifiers

Anna Krajuškina
Sven Nõmm
Department of Software Science,
Faculty of Information Technology
Tallinn University of Technology,
Akadeemia tee 15 a,
12618, Tallinn, Estonia
E-mail: sven.nomm@ttu.ee.

Aaro Toomela
School of Natural Sciences
and Health,
Tallinn University
Narva mnt. 25, 10120,
Tallinn, Estonia
E-mail: aaro.toomela@tlu.ee

Kadri Medijainen
Eveli Tamm
Institute of Sport Sciences
and Physiotherapy
University of Tartu
Puusepa 8, Tartu 51014, Estonia
E-mail: kadri.medijainen@ut.ee

Martti Vaske
Dan Uvarov
Department of Neurology
and Neurosurgery
University of Tartu
Puusepa 8, Tartu 51014, Estonia

Hedi Kähär
Marita Nugis
Institute of Sport Sciences
and Physiotherapy
University of Tartu
Puusepa 8, Tartu 51014, Estonia

Pille Taba
Department of Neurology
and Neurosurgery
University of Tartu
Puusepa 8, Tartu 51014, Estonia
E-mail: pille.taba@kliinikum.ee

***Abstract*—Motion mass based approach is adopted in this paper to describe gait movements of patients with Parkinson's disease. In spite of the recent advances in the areas of motion capture and motion analysis, medical community remains sceptical about applying computer aided systems to support modelling and diagnostics of the Parkinsons disease. The set of measurable parameters is limited by the time, step lengths and in some cases angles between the limbs. To complement these data, motion mass parameters describing amount and smoothness of the motions, are computed for each step. Decision tree classifiers a trained to distinguish between the gait motions of patients with diagnosed Parkinsons disease and healthy individuals of the same age. Moreover it is demonstrated that inclusion of the time does not increase model quality.**

I. Introduction

The present paper is devoted to the problem of Parkinson's disease recognition by gait analysis of the forward walking movement during the modified Up-and-Go test. Up to a very recent time medical community has remained sceptical about applicability of computer assisted systems for diagnostics and modelling of neurological disease. For the more advanced and precise systems main obstacle is lengthy and complicated set-up procedure which makes them nearly impossible to use on everyday basis. For the simpler systems, like Microsoft Kinect, main argument is their imprecision especially in capturing motions of the lower limbs [1]. In spite of these drawbacks Kinect and systems alike attract more and more attention as the potential tools to be used on everyday basis, for example in combination with neuro-fuzzy classifier [2] or purely with artificial neural networks [3]. Parkinsons disease (PD) is a neurogenerative disease, which affects human motor functions, resulting in slowing of movements and muscle rigidity [4]. Not only the precise diagnostics is a complicated task, at the present moment there is no known cure against PD [5]. Being early diagnosed and properly monitored the symptoms of PD may be improved, allowing individual to continue daily activities. These three factors motivate necessity of current research. Also development of the computer support systems for diagnostics, modelling and analysis of PD is constantly developing field of research. Large variety of methods that invariant of hardware setting include [6], applies wavelet transform to extract gait related features to be used in PD classification. Spatial Temporal features are analysed in [7]. Gait variability of PD patients is studied in [8]. Gait and tremor are analysed for PD diagnosis in [9]. Approach to support classification by visualising the data is proposed in [10]. Moreover in spite of the imprecision pointed out in [1] there are number of contributions demonstrating that precision of the sensor allows to distinguish individuals affected by the Parkinson's disease from the similar age group of healthy individuals (controls) [11]. Main distinctive feature of this paper is the way to compute features describing individual steps. Motion mass parameters were initially proposed in [12], [13] to describe learning of a new motor activity and later adopted for the Up-and-Go test used to evaluate motor performance of individuals with diagnosed PD [11]. In this paper the set of motion mass parameters will be used to describe individual steps. Also Decision tree classifiers are chosen among the other machine learning techniques due to their interpretability. Organisation of the paper follows the following template. Section II describes experimental setting and presents formal problem statements. Adaptation of motion mass parameters for the case of gait analysis is discussed

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00630

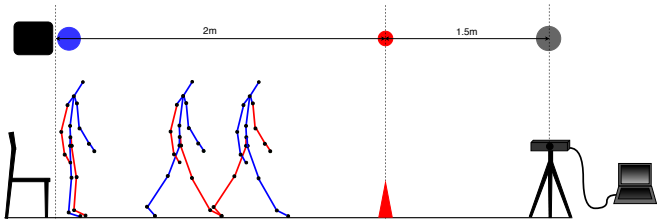

Fig. 1. Experimental setting.

in section II-B. Applicability of the different classification techniques to distinguish between the PD patients and controls is discussed in section III. Two final sections are focused to the short discussion of the achieved results and concluding remarks.

II. Problem statement, experimental setting and methodology

Formally research problem may be stated as follows. Design a method to objectively distinguish between PD patients and controls based on the forward walking gait analysis. This problem naturally splits into three following sub problems.

1) Precise the setting and capture the gait during forward walking motion.
2) Choose the set of features possessing the highest discriminative power.
3) Choose and train machine learning technique(s).

A. Experimental setting

Solution of the first sub problem partially based on the common medical practice to use Up-and-Go test and partially belongs to the hardware selection and software development. Up-and-Go test is widely used in medicine to assess motor condition of the patient. In its classical version the setting of the test requires a chair, approximately three meters of space to walk and chronometer. The measurable parameter in its classical setting is time. The test was divided in seven phases to be repeated three times. It starts with the patient sitting on the chair (first phase). Upon the practitioners command, the patient should stand up (second phase), **walk forward** towards a marker positioned on the floor (third phase), turn around (fourth phase), walk back to the chair (fifth phase), turn around (sixth phase) and seat on the chair (seventh phase). In such setting it is frequently referred as Timed Up-and-Go (TUG) test. Note there are many variations of test. Results reported below are based on the gait captured during forward walking motion (the third phase) of the Up- and-Go test. Walking gait may be observed only during walking forward (third phase) and walking back (fifth phase). Based on the restrictions of the hardware (Kinect may not acquire correctly walking backward motion), third phase was chosen to perform the analysis. Positions of the Kinect sensor, turning point marker and a chair are depicted in Fig.1.1. Particular distances are dictated by the fact that Kinect is able to generate model of a human with the range between 1.5 and 4 meters. Motion capture were performed with the Kinect sensor. Kinect returns the skeleton-like model of the human body described by twenty (twenty-five for the newer version) points. The choice of the sensor is justified by the fact that it does not require procedures inherent to the more sophisticated systems. Such procedures are attaching reflective markers to the human body and/or modelling the body in specialised software. Obviously, for daily practice such time consuming procedures are most likely unacceptable. For the case of patients with PD or other neurological disease such procedures may be simply impossible to perform. Therefore an effort is made in direction of using simpler devices. Note, that while within the frameworks of present research Kinect sensor is used, proposed technique is agnostic with respect to the hardware and may be easily adopted to any similar motion capture system. The final element of the hardware setting is an averagely powered laptop (i5 processor and 8GB of RAM) running the software specially developed to record human movements during Up-and-Go tests. The software allows control over the motion capture procedure and distinguishing different phases of the test. Recording of the exercise or the phase of the exercise is represented in the form of numeric matrix where rows represent observation points (frames) and columns represent 3D coordinates of each recorded point. Together with frame number and time stamp there are 62 columns for older version of the sensor and 77 columns for a newer version, whereas the number of rows depend on the time duration of the exercise. Typically recording of entire test takes between 1.5 and 3 megabytes. All the following steps of processing, feature extracting and modelling were performed separately either in MATLAB environment or programming in Python. Totally 40 individuals participated in the trials, 20 are patients with diagnosed PD and 20 age and sex matched controls. The base parameters describing the participants are: average age is 69 years, average height 169.4 cm and average wight 79.4 kg. All the trials were conducted in accordance to the law, personal data protection regulations and with the permission by the ethics committee.

B. Methodology

The second sub problem, identified above, may be split into two. The first one is feature extraction (construction) from the test recordings. And the second one is actual feature selection. The feature construction is based on the method of computing Motion Mass parameters for each step of the forward walking motion. Motion Mass parameters were initially introduced in [12] and [13] as the measures of amount and smoothness of the motion. They were proposed to model learning of a new motor activity [12] then tuple of four parameters was proposed. Later in [11] the number of parameters grew to eight. For the set of joints $J = \{j_1, \dots j_n\}$ trajectory mass L_J, acceleration mass A_J, and combined Euclidean distance E_J are defined as follows:

$$A_J = \sum_{j \in J} \sum_{t=1}^{t_m} a_{j_t}$$

$$L_J = \sum_{j \in J} l_j$$

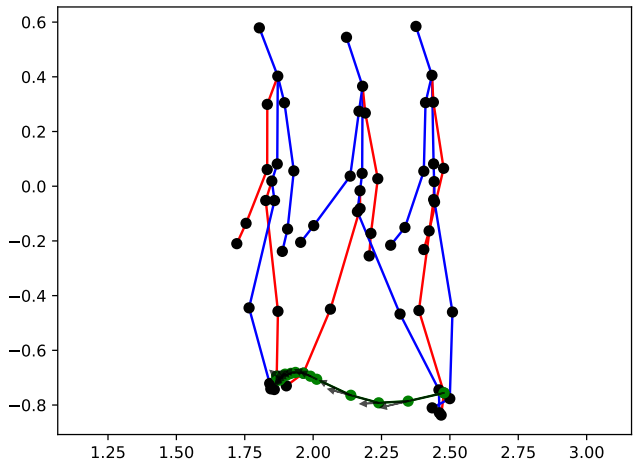

Fig. 2. One step.

$$E_J = \sum_{j \in J} s_j$$

where a_{j_t} is the acceleration of the joint j observed at time t, l_j is the length of trajectory followed by the joint j during the action (motion) and s_j is the length of the straight line segment connecting positions of the joint j in the beginning and in the ending of the experiment. The fourth parameter was the time length of the action [12]. In order to make the paper self-sufficient let us explain the meaning of this parameters on the example of making a single step. Let joints set of the interest consists of the points describing positions of hips, knees, ankles and feet. The step starts when the distance between the feet is minimal. Without loss of generality let us suppose that right feet goes ahead first. Once the right leg touches the ground the weight is transferred to it allowing left leg to be lifted off the ground and moved forward. Once the distance between the feet is minimal, next step starts. Fig. 2 depicts all the elements necessary to compute motion mass parameters for the left ankle during one step. Red colour represents left leg and blue colour -right leg. Black line with green dots represent the trajectory of left ankle. Arrows represent the directions of the tangent vectors to its trajectory at each observation point. In order to compute acceleration- and velocity masses one have to compute accelerations and velocities for each observation point in the direction of tangent vector. Velocity masses introduced in [11]. Also ratios of acceleration- and velocity- masses to the combined Euclidean distance were introduced in [11].

III. Feature Selection and Classifier Development

In the classical setting of Up-and-Go test time is the only one measurable parameter. One may consider either duration of entire test or duration of the forward walking (third phase of the test). It is easy to see, that its discriminative power to distinguish between PD patients and controls is far from being perfect. Fig. 3 depicts the distributions of average test durations observed for PD patients (red dots) and controls

TABLE I
Distinguishability and Discriminative Power of Time Duration.

	p-value	Fisher score
Overall Time	0.001634	0.329513
Walking Forward Time	0.004709	0.253020
Step Time	0.453093	0.015126

(blue dots). One may easily observe the overlapping between

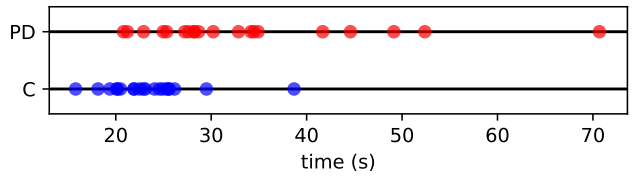

Fig. 3. Average test duration.

the intervals containing time duration of PD patients and controls. Average time durations of walking forward phase of the test see Fig.4 are similar to those of entire test. Finally

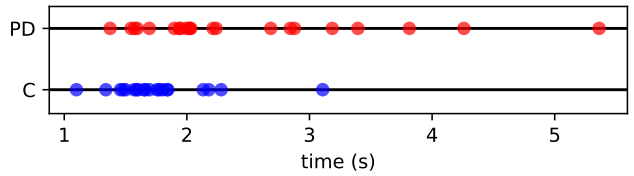

Fig. 4. Average duration of walking forward phase.

average duration of one step observed during the working forward phase are depicted by Fig. 5. It is easy to see that in this case time would not allow to distinguish between the PD patients and controls. More formally, results of the

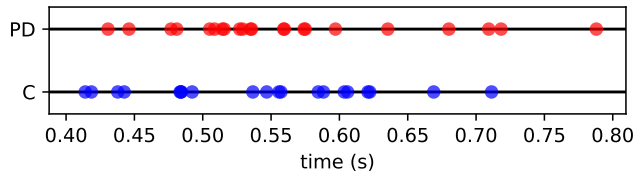

Fig. 5. Average step duration.

Welch's test and Fisher's score values for time duration are presented in Tab.I The same test demonstrates that there is a subset of joints such that mean values of at least one motion mass parameter differ significantly between PD patients and controls. Totally 48 differ significantly between two groups. Therefore, proposed experimental setting is sensitive enough to capture the difference between the motions of PD patients and controls. Based on the Fisher's scores in Tab.III, the following parameters were selected as possible candidates to build the classifier. In the table III first column contains name of the joint, second column denotes parameter name (E - Euclidean

TABLE II
p - VALUES OF THE DISTINGUISHABLE MOTION MASS PARAMETERS.

Joint Type	E	Lm	Vm
AnkleLeft	0.000038	0.110393	0.103945
AnkleRight	0.000040	0.122571	0.148019
ElbowLeft	0.000572	0.011432	0.008152
ElbowRight	0.000010	0.000166	0.000115
FootLeft	0.000050	0.098695	0.099398
FootRight	0.000040	0.141181	0.186130
HandLeft	0.012222	0.168180	0.171630
HandRight	0.000044	0.017727	0.015840
Head	0.000046	0.001897	0.001258
HipLeft	0.000025	0.016621	0.010584
HipRight	0.000008	0.012844	0.007028
KneeLeft	0.000050	0.004637	0.003581
KneeRight	0.000007	0.010762	0.007537
ShoulderLeft	0.000102	0.001018	0.000659
ShoulderRight	0.000008	0.000158	0.000112
SpineBase	0.000014	0.011382	0.006396
SpineMid	0.000014	0.012349	0.006631
SpineShoulder	0.000020	0.000190	0.000110
WristLeft	0.005845	0.059722	0.056430
WristRight	0.000030	0.003059	0.002610

TABLE III
HIGHEST FISHER'S SCORES OF THE MOTION MASS PARAMETERS.

Joint Type	MM	Fisher Score
KneeRight	E	0.656636
ShoulderRight	E	0.649326
HipRight	E	0.647051
ElbowRight	E	0.627928
HipLeft	E	0.561201
WristRight	E	0.544609
AnkleLeft	E	0.525724
AnkleRight	E	0.524872
KneeLeft	E	0.506691
ShoulderLeft	E	0.460907
ShoulderRight	Vm	0.449920
ElbowRight	Vm	0.449093
ElbowLeft	E	0.346938
ShoulderLeft	Vm	0.335927
WristLeft	E	0.206193
ElbowLeft	Vm	0.187964
KneeRight	Lm	0.173483
HandRight	Lm	0.153361
FootLeft	Lm	0.068773

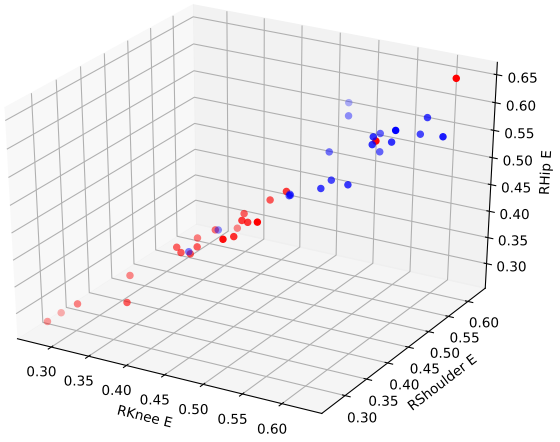

Fig. 6. Scatter plot with three most informative features.

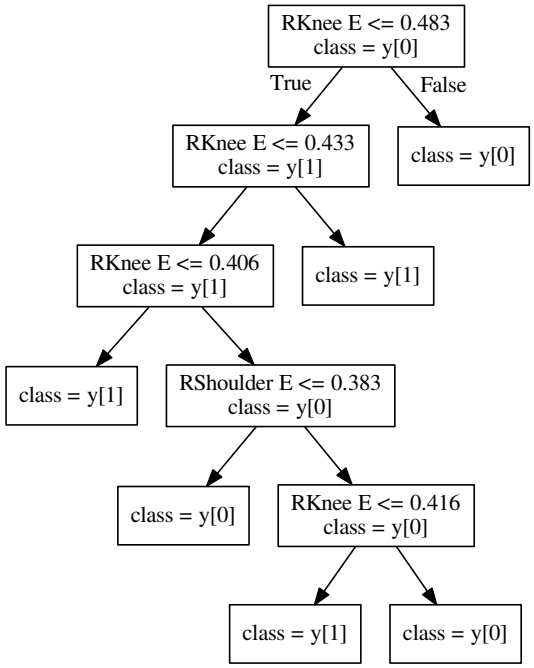

Fig. 7. Decision tree classifier for the initial feature selection.

distance, V_m velocity mass, L_m - trajectory mass.) and third column contains the value of the Fisher's score [14]. The parameters with the largest Fisher's score are right knee, right elbow, right hip and right shoulder. Observing scatter plot for three best parameters give an idea about distinguishing between PD patients and controls see Fig. 6. Taking into account relatively small sample size k - fold cross validation technique will be used to compute the accuracy of trained models. Choosing just two parameters, with highest Fisher's score as predictors lead the following decision tree classifier 7. Based on the 6- fold cross validation the accuracy of this predictor is 0.71. Corresponding confusion matrix is presented by Tab.IV, where (C) denotes controls and (PD) denotes patients with PD. Then duration of the experiment was added as third predictor. No difference in accuracy or confusion matrix was observed. Choosing the variables purely on the Fisher's score allows to achieve accuracy range $0.71 - 0.85$ depending on the number of predictors and parameters of the tree. Such approach may be acceptable form the machine learning viewpoint, but it eliminates interpretability of decision tree classifiers. Only right-side joints could not describe impairments caused by PD. The common practice in analysis of motor impairments caused by neurological diseases [5] is to consider either symmetrical or asymmetrical set of joints. For the symmetrical sets of joints corresponding features are: combined Euclidean distance of right knee, right hip ,right shoulder and combined Euclidean

TABLE IV
CONFUSION MATRIX FOR THE CLASSIFIER 7.

	Predicted (C)	Predicted (PD)
Actual(C)	3	2
Actual(PD)	0	2

TABLE V
CONFUSION MATRIX FOR THE CLASSIFIER. 9

	Predicted (C)	Predicted (PD)
Actual(C)	4	1
Actual(PD)	0	2

distance of symmetric left joints. Represented by these features separability of PD patients and controls may be observed in Fig.8. Corresponding classifier is depicted in Fig. 9 While for

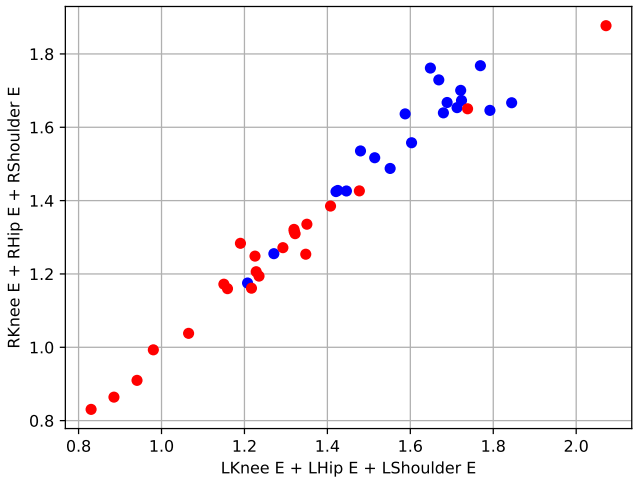

Fig. 8. Separability for the symmetric choice of features

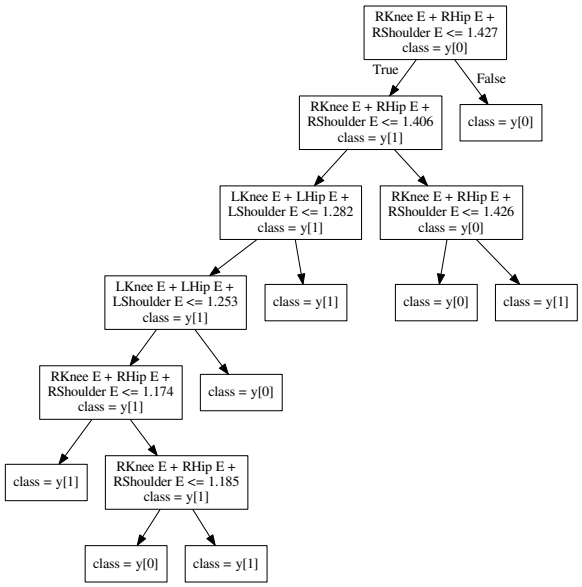

Fig. 9. Decision tree classifier for the symmetric feature selection.

some predictors Fisher's are lower, compared to the previous case, the model accuracy did not change much and remains in the range of $0.75 - 0.85$ (depending on the fold in cross validation). Confusion matrix corresponding for these case is presented in Tab. V Adding time duration of the test did not produce any observable effect. For the case of asymmetrical set the following features are selected combined Euclidean distance for the right shoulder, right elbow, right wrist, left hip, left knee and left ankle, and combined Euclidean distance of the symmetric joints. Represented by these two features separability of the PD patients and controls is depicted by Fig.10. Corresponding classifier is depicted in Fig. 11 In this

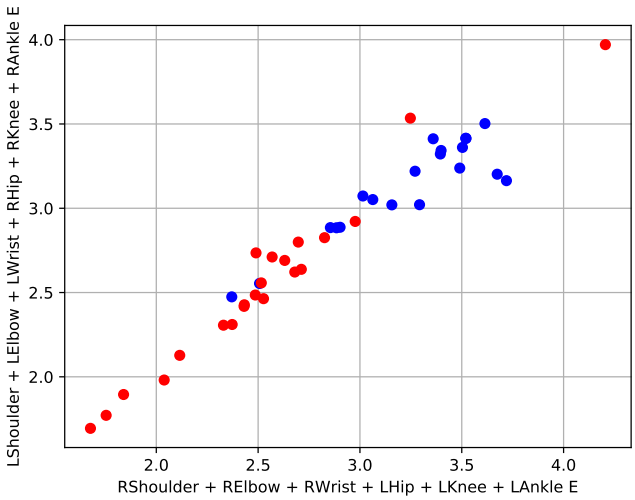

Fig. 10. Separability for the asymmetric choice of features.

case the results are the same as in previous case. Also inclusion of time as a predictor did not affect the quality of the model. Among kinematic features the velocity mass of right elbow possesses highest discriminative power. Taking its value into account lead following decision tree classifier 12. Confusion matrix and accuracy of this classifier are in pair with other models considered. An interesting observation is that velocity mass parameters for the left elbow is the second condition to be verified which indicate importance of kinematic parameters in diagnosing PD.

IV. DISCUSSION

The structures of the decision tree classifiers confirm the importance of knee motions in distinguishing PD patients from controls. It is demonstrated that features containing properties of knee motions are present in a higher nodes of the graphs describing the classifiers. The parameters describing the motions of the shoulders and elbows also possess certain power to distinguish between the movements of PD patients and controls. An important observation is that overall time of the test possess relatively low discriminative power. This was

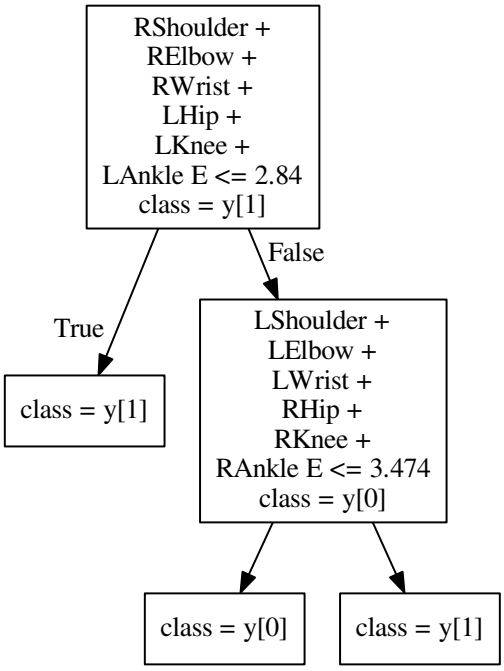

Fig. 11. Decision tree classifier for the symmetric feature selection.

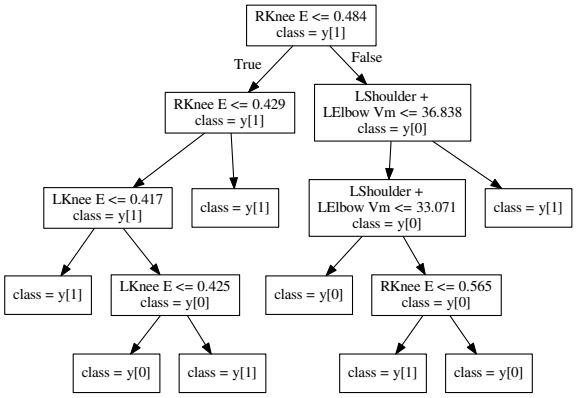

Fig. 12. Decision tree classifier for the feature set including velocity mass.

demonstrated by relatively low values of the Fisher's score and by the fact that including time duration of the test as the predictor did not improved the model. One may argue that overall precision of the proposed techniques is lesser compared to those archived with support vector machines, neural networks or boosted methods. While precision of the single decision tree is slightly lesser compared to other techniques. It is compensated by the traceability of the decision and its interpretability.

V. Conclusions

Motion mass base method is adopted in this paper to distinguish between the PD patients and group of control individuals of the similar age on the basis of captured gait movements. Analysis of the structure of trained decision trees has clearly demonstrate that motions of the knees, shoulders, and elbows differ significantly between the groups of PD patients and controls. Also it demonstrates that for the analysis of gait motions time does not possess any discriminative power. These points provide interesting directions for the further research of motions affected by PD.

References

[1] X. Xu and R. W. McGorry, "The validity of the first and second generation microsoft kinect for identifying joint center locations during static postures," *Applied Ergonomics*, vol. 49, pp. 47 – 54, 2015.

[2] O. Geman, "Nonlinear dynamics, artificial neural networks and neuro-fuzzy classifier for automatic assessing of tremor severity," in *2013 E-Health and Bioengineering Conference (EHB)*, Nov 2013, pp. 1–4.

[3] O. F. Ince, I. F. Ince, J. S. Park, and J. K. Song, "Gait analysis and identification based on joint information using rgb-depth camera," in *2017 14th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, June 2017, pp. 561–563.

[4] A. Berardelli, G. Wenning, A. Antonini, D. Berg, B. Bloem, V. Bonifati, D. Brooks, D. Burn, C. Colosimo, A. Fanciulli *et al.*, "Efns/mds-es recommendations for the diagnosis of parkinson's disease," *European Journal of Neurology*, vol. 20, no. 1, pp. 16–34, 2013.

[5] L. V. Kalia and A. E. Lang, "Parkinson's disease," *The Lancet*, vol. 386, no. 9996, pp. 896 – 912, 2015.

[6] M. S. Baby, A. J. Saji, and C. S. Kumar, "Parkinsons disease classification using wavelet transform based feature extraction of gait data," in *2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT)*, April 2017, pp. 1–6.

[7] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, "Classification of parkinson's disease gait using spatial-temporal gait features," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1794–1802, Nov 2015.

[8] L. Ota, H. Uchitomi, S. Orimo, and Y. Miyake, "Classification of parkinson's disease patients' gait variability," in *2012 IEEE/SICE International Symposium on System Integration (SII)*, Dec 2012, pp. 343–348.

[9] E. Abdulhay, N. Arunkumar, K. Narasimhan, E. Vellaiappan, and V. Venkatraman, "Gait and tremor investigation using machine learning techniques for the diagnosis of parkinson disease," *Future Generation Computer Systems*, vol. 83, pp. 366 – 373, 2018.

[10] U. K. Pun, H. Gu, Z. Dong, and N. S. Artan, "Classification and visualization tool for gait analysis of parkinson's disease," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2016, pp. 2407–2410.

[11] S. Nõmm, A. Toomela, M. Vaske, D. Uvarov, and P.Taba, "An alternative approach to distinguish movements of parkinson disease patients," in *The 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*, August 2016, p. Accepted.

[12] S. Nõmm and A. Toomela, "An alternative approach to measure quantity and smoothness of the human limb motions," *Estonian Journal of Engineering*, vol. 19, no. 4, p. 298308, 2013.

[13] S. Nomm, A. Toomela, and J. Borushko, "Alternative approach to model changes of human motor functions," in *Modelling Symposium (EMS), 2013 European*, Nov 2013, pp. 169–174.

[14] C. Aggarwal, *Data Mining: The Textbook*. Springer International Publishing, 2015.