Research Article

Feasibility of Using Dynamic Time Warping to Measure Motor States in Parkinson's Disease

Somayeh Aghanavesi [ID],[1] Hasan Fleyeh,[1] and Mark Dougherty[2]

[1]*School of Technology and Business Studies, Dalarna University, Falun 78188, Sweden*
[2]*School of Information Technology, Halmstad University, Halmstad 30250, Sweden*

Correspondence should be addressed to Somayeh Aghanavesi; saa@du.se

The aim of this paper is to investigate the feasibility of using the Dynamic Time Warping (DTW) method to measure motor states in advanced Parkinson's disease (PD). Data were collected from 19 PD patients who experimented leg agility motor tests with motion sensors on their ankles once before and multiple times after an administration of 150% of their normal daily dose of medication. Experiments of 22 healthy controls were included. Three movement disorder specialists rated the motor states of the patients according to Treatment Response Scale (TRS) using recorded videos of the experiments. A DTW-based motor state distance score (DDS) was constructed using the acceleration and gyroscope signals collected during leg agility motor tests. Mean DDS showed similar trends to mean TRS scores across the test occasions. Mean DDS was able to differentiate between PD patients at Off and On motor states. DDS was able to classify the motor state changes with good accuracy (82%). The PD patients who showed more response to medication were selected using the TRS scale, and the most related DTW-based features to their TRS scores were investigated. There were individual DTW-based features identified for each patient. In conclusion, the DTW method can provide information about motor states of advanced PD patients which can be used in the development of methods for automatic motor scoring of PD.

1. Introduction

Parkinson's disease (PD) is a chronic progressive disease characterized by motor and nonmotor symptoms which affect the quality of life of patients [1, 2]. Motor symptoms are phenomena such as tremor, rigidity, and bradykinesia. In the advanced stages of PD, motor states of PD patients fluctuate between three states: "On," "Off," and "On with dyskinesia." "Off" state is when insufficient medication effect causes patients to experience Parkinsonism symptoms, while during "On" state, relief of symptoms occurs due to enough medication effect. "On with dyskinesia" is a state where patients experience involuntary movements due to excessive amount of medication. Dyskinesia can occur in response to both increasing and also decreasing concentrations followed by a sudden "Off" state [3–5].

Treatment of PD is guided by clinical examination using rating scales. The most common one is the unified PD rating scale (UPDRS) [6]. UPDRS section III is used for clinical scoring of the motor symptoms. Another scale used for assessing the motor states in PD is the Treatment Response Scale (TRS) [7]. However, the rating scales are subject to intra- and interobserver variabilities; they are not able to capture variations in symptoms continuously. Patients require physical visits to the clinic for their motor symptoms to be rated. The clinical ratings are done infrequently, not including events at home, thus do not provide a full picture of the patients' condition. This is problematic given that their motor states can fluctuate from time to time during the day [8].

The development of methods for continuous and automatic PD motor state assessment has been done in various studies [9–11]. Different data- and knowledge-driven methods have been used to extract information for the development of such tools. Clinical assessment of PD symptoms is done visually by clinical experts observing the speed, rhythm, and extent of the movements according to rating scales. PD
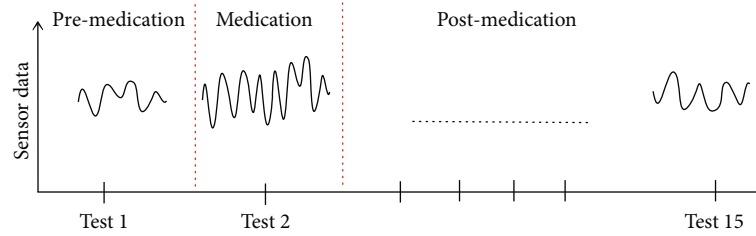
FIGURE 1: Outline of the study showing where DTW method is used to measure the distances between leg agility signals.

symptoms can manifest in various directions and degrees depending on patients' motor states. In other words, when PD patients are Off, On, or dyskinetic, they may experiment the similar tasks but at different amplitudes and frequencies. Therefore, the recorded signals of a repeated task experimented at different motor states can be dissimilar and they can vary in length and frequency. This can be reflected when measuring the similarity of the signals recorded from similar tasks but experimented at different motor states. An approach to measuring the similarity of the signals with different lengths and frequencies is the Dynamic Time Warping (DTW) method.

DTW was first introduced in the 1960s and initially became popular in the context of speech recognition [12] and time series data mining. It is a technique used to find the optimal alignment between two time-dependent series [13]. With applications in finance, DTW was used to find similar historical subsequences and predictions were made from the mapping of the most similar subsequences [14].

Previous studies [15–18] employed the DTW method to segment the signals, part of which matched well with a predefined reference pattern. Ghassemi et al. [15] employed two variations of this method in different experiments of PD patients walking for gait segmentation. The experiment included a straight walk, a heterogeneous assessment paradigm, and walking including turns were examined with the method. They achieved high accuracy using this method for segmenting all types of gait sequences. Barth et al. [16] segmented gait signals from daily life activities into single steps for further computation of step parameters. They used gyroscopes attached to sports shoes and used the DTW method to recognize the beginning and end of steps during the walking activity. They achieved a 97% recognition rate of steps. Barth [17] developed a mobile system to rate gait impairment. The authors used a multidimensional subsequence DTW method to extract single strides from gait tests and free walking. They used a predefined stride template to find a matching pattern in PD patients' walking signals. They achieved excellent recognition rates for both experiments.

DTW was also used for matching voice samples in PD. Vikas Sharma [18] used DTW to distinguish PD patients from healthy controls and achieved a high DTW-based matching percentage of 80% for PD patients vs. 72% for healthy controls. In addition, Adame et al. [19] used the DTW method in timed up and go (TUG) test for distinguishing between healthy controls, early PD, and advanced PD patients. They attached a wearable inertial sensor unit to the lower back of PD patients and healthy controls and

instructed them to experiment with the TUG test. Results from using this method showed that the state transitions of sit to stand and turning for healthy controls and early PD patients vs. advanced PD patients were significantly different. With assessment of PD symptoms, a study is aimed at proposing a method based on DTW for monitoring the changes in gait trajectories. They used motion sensors to record hip and knee inclination (pitch) of four PD patients and six healthy controls during walking. They compared the average distances between healthy subject and PD patients and also the distance between the experiments PD patients had before and after medication. It was discovered that this method can be used for discriminating a healthy gait from an impaired one. In addition, it was found that this method is useful for monitoring the changes in gait pattern of PD patients before and after medication. However, the number of subjects was few, the significance of the scores from results was not investigated, and the raw signals of the walking experiments were not used in DTW.

In spite of that, since there is a high risk of falling during walking in older PD patients, this study focuses on using leg agility test as a better choice for evaluation of the motor symptoms. Previous researches have shown leg agility contains important information about disease severity [20, 21]. To measure the motor states of the PD patients in a previous study [10], the authors have used different techniques for extracting 24 features to be included in developing the methods for assessment of PD motor states. However, the DTW method was not included as part of the features since the feasibility of using this method for measuring the motor states was unclear. Besides, to our knowledge, DTW has not been examined for its usefulness for measuring the motor states in PD using the raw signals of the experiments. This method provides a measure of similarity of the signals regardless of their length and frequency. Feasibility of using this method for measuring the motor states can provide information to be used for the development of methods for automatic motor scoring of PD.

This study is aimed at investigating the feasibility and the extent of using a distance measure calculated by the DTW method to assess the motor states of advanced PD patients over a course of UPDRS-based leg agility motor tests. An outline of the proposed approach is shown in Figure 1.

For this purpose, extracting information from DTW-based distances between 3D motion sensor signals of experiments recorded at different motor states, a DTW-based motor state distance score (denoted as DDS) was developed and its properties was examined. To investigate whether

DDS can detect the motor states similar to how it was assessed by TRS, the mean of DDS vs. the mean of TRS scores across the tests was examined. The trend of DDS for healthy subjects was included. Further, the power of the method in quantifying the motor states was investigated for the most responsive PD patients to medication. Patients were selected through their TRS scores, and the most informative DTW-based measures/features about their motor states were identified and evaluated. To find out whether DDS is significantly different between groups of PD patients at different motor states as scored by TRS, the mean DDS across groups of patients and healthy subjects was investigated. The feasibility of using the developed DTW-based scores for classifying the motor state changes was investigated.

2. Materials and Methods

2.1. Participants. In this study, 19 advanced PD patients experiencing motor fluctuations were recruited to an observational study where they were given a single dose of medicine in a hospital in Uppsala, Sweden [22]. 22 healthy controls experimented the same tests for up to 8 times. The study was carried out in one center and was approved by the regional ethical review board in Uppsala, Sweden (reference number: 2015/100). The patient characteristics are shown in Table 1.

2.2. Experimental Setup. Subjects put on their ankles a 3-axial accelerometer and gyroscopes with a sampling rate of 102.4 Hz, accelerometer range of +/-16 g, and gyroscope range of +/-2000 dps. They experimented the leg agility test by sitting on a straight back chair, placing both feet on the floor, and tapping one foot at a time (first right foot and then left foot) for 10 times and as fast as possible. The test is defined in item 26 of UPDRS III where the agility of the foot tapping of PD patients is evaluated for assessment of the severity of the motor states. The sensor data of all time points (X, Y, and Z axes of accelerometers and gyroscopes) were collected and saved. The foot tapping signals were segmented by identifying the movements in the acceleration data in the Y-axis. The details of measuring the experiments by motion sensors, data preprocessing, and segmentation were described in a previous study [10]. The extracted segments consisting of 3D accelerometer and gyroscope data were used to calculate the features as described in the next section.

For PD patients, the leg agility motor tests were experimented at different time intervals for up to 15 trials. The first experiment was within 50 min before taking the dose. The second experiment was right after administration of 150% of their normal daily dose (0 min). Administration of the levodopa dose was to follow the individual response of the patients to the dose. It was to explore motor state change from Off (within -50 min) to good mobility and/or dyskinetic state (expected within 20-60 min) and back to Off state. Written informed consent was given. The next tests were approximately at 20, 40, 60, 80, 110, 140, 170, 200, 230, 260, 290, 320, and 350 min after the dose administration. Healthy controls experimented the test up to 8 times.

The experiments were video recorded. The segmented data were synchronized with the videos. The videos were shown in a randomized order to three movement disorder specialists for rating the experiments of the patients regardless of the time points the medications were taken. The clinical experts rated the overall mobility of patients, using the Treatment Response Scale (TRS) [7] ranging from -3 (very "Off," severe Parkinsonism) to 0 (On, normal mobility) to 3 (severe dyskinesia, severe choreatic dyskinesia). They rated severity of dyskinesia on a scale of 0 (no dyskinesia) to 4 (severe dyskinesia) [6], and some items of UPDRS section III including UPDRS #27 (Arising from chair), UPDRS #29 (gait), and UPDRS #31 (body bradykinesia) each of them rated on a scale from 0 (normal) to 4 (severely impaired).

In addition to leg agility tests, experiment of the PD patients included walking and hand rotation. During the walking experiment, subjects performed a 2.5-meter walk. During the hand rotation experiment, subjects performed alternate pronating hands for 20 seconds. The application of DTW was done on walking and hand rotation data when the first test was set as a baseline signal for measuring the distances of all other tests to it and also when the distances of every consecutive test signals were measured. The best results were achieved when using the second approach on leg agility test data.

2.3. Extraction of Distance Measures by DTW. The different signals extracted from the foot tapping tests for PD subjects were displaced in time, and they had different lengths and frequencies. Calculating the Euclidian distance between each two data points from the two time series is not an ideal approach for finding the distance between two signals since it is not providing invariance to warping for time series that differ in length. Instead, in order to measure the distance between these two consecutive test time series, the DTW was used to dynamically adjust the metric in order to find better alignment between them [23]. The details of the DTW can be found in the study by Keogh and Ratanamahatana [24]. This algorithm can be described in two steps. First, it finds the distance between the two time series of (X_i) $(1 \le i \le n)$ and (Y_j) $(1 \le j \le m)$ resulting in a matrix D_{ij} with a dimension of $n \times m$, containing distances between X_i and Y_j. The distances within the matrix were then calculated by the sum of the distances between the two elements comprising the series of X_i and Y_j. Finally, the minimum of the neighbouring elements in matrix D_{ij} was calculated as [24]

$$D_{ij} = \left| X_i - Y_j \right| + \min \left\{ D_{i-1,j}, D_{i-1,j-1}, D_{i,j-1} \right\}. \quad (1)$$

Figure 2 is an example that illustrates the acceleration signal collected from the Y-axis during experimenting one-foot tap during the first interval (test 1, premedication, blue) together with that experimented at the second interval (test 2, at the time of medication, red). Figure 2(b) shows the original signals while Figure 2(c) shows that the two time series are warped. In Figure 2(d), the distance matrix containing the distances and the optimal path is shown. The figure illustrates only one tap trial for clarity.

TABLE 1: Characteristics of participants shown as mean (standard deviation), levodopa-carbidopa study dose as min-max.

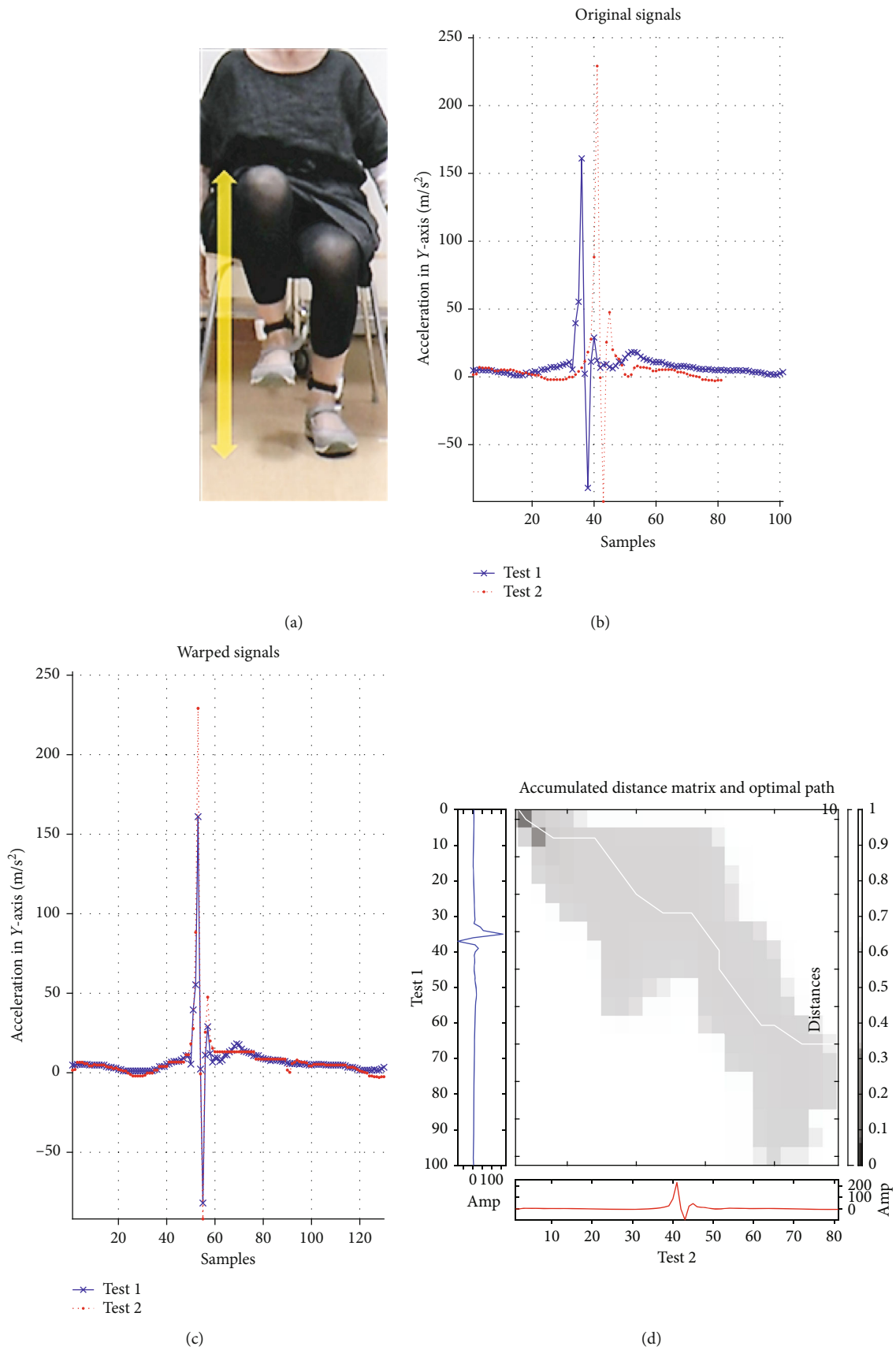|	Gender	Age (years)	Height (m)	Weight (kg)	Years with PD	Years on levodopa	Symptom duration (year)	Levodopa-carbidopa study dose (mg)	Affected side	Hoehn & Yahr	UPDRS IV
Patients	14 males, 5 females	71.4 (6.3)	1.75 (0.09)	75.4 (11)	9.7 (6.8)	9.5 (6.5)	12.2 (7.3)	110-410	9 right, 10 left	3.1 (0.8)	6.21 (3.13)
Healthy controls	16 males, 6 females	64.2 (7.4)	1.75 (0.1)	83.6 (13.8)	—	—	—	—			

(a)

(b)

(c)

(d)

FIGURE 2: (a) Leg agility performance along Y-axis. (b) Original signals captured by sensors for a single foot tapping. Prior test is blue line and later test is red line. (c) The signals are warped. (d) Accumulated distance matrix and optimal path.
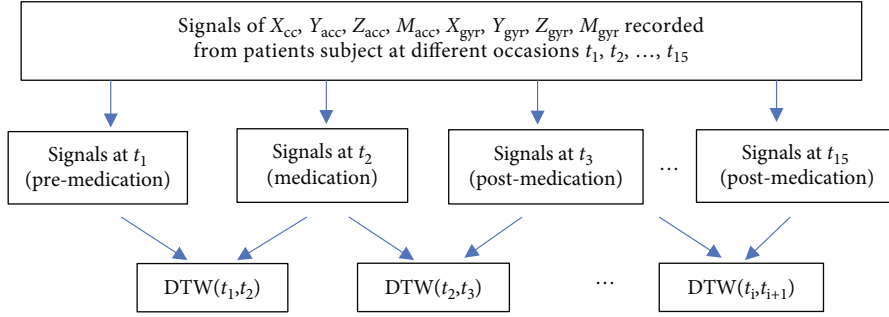
FIGURE 3: Illustration of calculating the distance measures and preparing the dataset.

Distance measures were calculated for full experiment of foot tapping.

2.4. Creating the Dataset. There were six signals recorded by motion sensors from the original foot tapping test each time this test was conducted. These signals are X_{acc}, Y_{acc}, and Z_{acc} to represent the acceleration of the foot tapping in the X, Y, and Z axes and X_{gyr}, Y_{gyr}, and Z_{gyr} to represent the gyroscopic rotation of the leg around the X, Y, and Z axes according to the right-hand rule, respectively. In addition to those signals, the magnitude of acceleration (M_{acc}) in m/s^2 and the magnitude of orientation (M_{gyr}) in °/s were calculated using

$$M_{acc} = \sqrt{X_{acc}^2 + Y_{acc}^2 + Z_{acc}^2}, \qquad (2)$$

$$M_{gyr} = \sqrt{X_{gyr}^2 + Y_{gyr}^2 + Z_{gyr}^2}. \qquad (3)$$

The distance measures were calculated between every two consecutive tests using the mentioned eight variables. Figure 3 presents the approach for preparing the dataset.

Let $X_{DTW(t_i,t_{i+1})}$, $Y_{DTW(t_i,t_{i+1})}$, and $Z_{DTW(t_i,t_{i+1})}$ be the DTW calculated for two signals of the same axis from two consecutive tests, then the following is valid for the acceleration as well as the gyroscope signals.

$$X_{DTW(t_i,t_{i+1})} = DTW\left(X_{t_i}, X_{t_{i+1}}\right), \qquad (4)$$

$$Y_{DTW(t_i,t_{i+1})} = DTW\left(Y_{t_i}, Y_{t_{i+1}}\right), \qquad (5)$$

$$Z_{DTW(t_i,t_{i+1})} = DTW\left(Z_{t_i}, Z_{t_{i+1}}\right). \qquad (6)$$

Equations (4), (5), and (6) are calculated for the acceleration values as well as when they are calculated for the gyroscopic X, Y, and Z values. Calculating the distances using individual axis signals and the magnitudes of acceleration and orientations was done to capture all deteriorations and differences in individual directions as well as in their overall weight. By using these signals, 16 features were calculated.

Two more features were added after calculating the mean of the magnitude of the acceleration and the gyroscope distance measures of the right and left foot. This was to examine if the distance between the mean of the acceleration and mean of the gyroscope signals from both feet can provide motor state information.

Since the leg movements during the foot tapping experiment were along the Y-axis and during different motor states, the acceleration of the foot tapping can differ in this direction; the mean of the distance measure in the Y-axis was calculated for right and left foot resulting in the last feature. Possible differences in directions of Y and Z are investigated for each foot separately. In the end, there were a total of 19 features.

An example of the dataset including the measures for one patient is presented in Table 2.

Referring to Table 2, the first row on top represents the two consecutive tests that the distance measure was calculated for. The values in each cell represent the distance measure calculated for those two respective tests and using the signals that are presented at the headers of each row. The signals are mainly in two groups. Eight signals for the right (R) and eight for the left foot (L). Also, the signals from each foot are in two groups of acceleration signals and orientation signals. Acc stands for acceleration and Gyr stands for gyroscope data. Both contain the signals at individual axes and their magnitudes (M). Three rows in Table 2 at the bottom represent the features for mean right and left foot acceleration in the Y-axis (Mean_RL_Acc_Y), the mean right and left foot magnitude acceleration (Mean_RL_M$_{acc}$), and the mean right and left foot magnitude orientation (Mean_RL_M$_{gyr}$), respectively. The total number of observations were 229 and 153 for patients and healthy subjects, respectively.

2.5. Statistical Analysis. The 19 calculated features were used to extract a motor state distance score (DDS). To identify which distance measures reflect more information about motor states as scored by TRS, a stepwise feature selection method with the fast-forward approach was first used where TRS was set as the response. Selection of the stepwise method was based on the recommendation of a previous study which identified it as a good choice for selecting the features and improving the performance of analytic methods [25]. Since there was a high agreement between the three clinical raters on the TRS scale (ICC = 0.82) as assessed in the previous study [10], this scale was used as the response to the selection method. This method was applied to a dataset containing all patients. To extract a score containing the most variability from all features, principal component analysis (PCA) was then applied to the selected features. PCA is a method to reduce the dimension of the feature space by identifying the

TABLE 2: Distance measures for one patient. R is the right foot and L is the left foot. The values correspond to the calculated distances between test occasions. Acc is acceleration. Gyr is gyroscope. M_{acc} is magnitude acceleration. M_{gyr} is magnitude gyroscope. Acc_Y is acceleration in the Y-axis.

Tests			1, 2	2, 3	3, 4	4, 5	5, 6	6, 7	7, 8	8, 9	9, 10	10, 11	11, 12	12, 13	13, 14	14, 15
R	Acc	X	0.2	0.01	0.2	18	10	5	0.9	14	10	2	5	0.2	5	3
		Y	1.9	1	5	0.2	0.2	60	47	25	5	0.2	0.2	0.2	0.2	3
		Z	0.8	0.01	0.01	1	2	4	1	1	0.6	1	2	0.2	3	0.3
		M_{acc}	135	17	3	438	88	511	3	2509	1599	0.2	18	0.8	169	177
	Gyr	X	51	0.7	189	4185	7763	631	6268	6917	2965	6	110	11	53	465
		Y	910	273	568	89	15115	33307	20243	3091	2269	506	0.9	200	635	1130
		Z	1	1	5	1	3	54	46	10	0.7	0.01	1	0.4	0.01	5
		M_{gyr}	465	241	415	2630	5677	9143	2825	243	5847	195	29	117	242	653
L	Acc	X	0.01	1	3	1	1	1	0.01	3	1	0.01	0.6	0.2	0.01	5
		Y	1	1	5	1	3	54	46	10	0.7	0.01	1	0.4	0.01	5
		Z	2.5	2	0.01	2	0.5	0.9	0.1	0.01	0.1	0.01	0.01	0.2	0.4	0.3
		M_{acc}	591	901	8660	1562	6206	12435	11843	2541	1410	529	1	251	287	0.1
	Gyr	X	1	0.3	1445	2715	841	878	3	844	137	0.01	3	45	6	182
		Y	135	669	2668	5831	1519	22	42839	2398	5387	125	1	539	313	869
		Z	18	1	105	49	17	93	34	0.2	23	0.3	1	0.5	9	3
		M_{gyr}	67	356	4698	197	8102	6602	6	4312	239	519	1	749	567	735
Mean_R_L		Acc_Y	1	1	5	0.7	1	57	47	17	2	0.1	0.7	0.3	0.1	4
		M_{acc}	10	1	55	25	10	74	40	5	12	0.2	1	0.5	4	4
		M_{gyr}	266	299	2556	1414	6889	7872	1416	2277	3043	357	15	433	404	694

most and least important variables [26]. For this, a matrix summarizing the relation between variables is built. Then, the matrix is divided to two separate components: direction and magnitude. Eigenvalues as a measure of covariance in data are the coefficients attached to eigenvectors providing the axis magnitude. Ranking the eigenvectors in order of their eigenvalues, highest to lowest, the order of the significant principal components is achieved. Two principal component (PC) scores with eigenvalues higher than 1 and containing 0.53% of the total variation were retained. The scores of the first PC in the vector were ranked and then rescaled to the range of the TRS scale (-3 to +2) per patient in the dataset resulting in DDS. Mean DDS between each test occasion (time points) ± 0.95 confidence intervals of the mean for each test occasion was investigated. This was to investigate the trend of mean DDS in capturing the motor states over the course of the test occasions vs. the trend of mean TRS scores. For healthy controls, the features were extracted. Since stepwise cannot be applied on a group with one value as target, the first principal component was used for the analysis.

Further, the extent that the individual DTW-based features matched the TRS scores of individual patients was investigated. For this, seven patients who showed a response to medication (from Off to On and back to Off) were selected. To identify the most related DTW-based feature to each patient, their features were used as input to stepwise selection method where their TRS scores were set as response. Moreover, the correlation coefficient of the identified most important DTW-based feature to their TRS score was examined.

The power of the DDS for separating the different motor states as scores by TRS was investigated using ANOVA. For this, the mean DDS between the groups of PD patients at different motor states was assessed.

To assess the differences in DDS results between right and left legs of the PD patients mostly affected on the right or left side, t-tests were performed on the first PCs of the selected features for each individual leg.

Using DTW, the distance between two consecutive signals that were recorded at different motor states was calculated. Therefore, the power of DDS in identifying the changes between the motor states of the patients as scored by TRS was examined. The amount of increase/decrease in TRS scores from one test to the next is calculated per patient. For this, the primary TRS score is subtracted from the secondary TRS score, e.g., if the first test TRS score was +1 and the next test score was +2, then the amount of change is calculated as $+2 - (+1) = -1$. Corresponding all DDSs with the calculated measures of changes in TRS, there were 211 observations left since the TRS difference could not be measured for the last test. Moreover, there were 12 missing values which were excluded, resulting in 199 observations in the dataset. Several classifiers were examined, and this study proposes the two best performing ones:

(i) Support Vector Machines (SVM) performing the classification by finding the hyperplane that maximizes the margin between the classes [27]. SVM use a kernel trick technique to transform the data to find the optimal boundary between the possible outputs. In this case, a common kernel function, the radial basis function kernel, was used with a gamma value of 0.0 and an epsilon value of 0.001

(ii) Decision Tree (DT) is an intuitive model making decisions based on the sequence of evaluations made to the feature values [28]. It makes the model in a tree structure by breaking the data into smaller subsections where leaf nodes represent the decision or classification. In this case, the number of training examples (batch size) was set as 100, where the minimum total weight of the instances in a leaf was set as 2 and no restriction was set for the depth of the tree 10-fold crossvalidation was used for both classifiers where the whole data set was divided into 10 sets; 9 sets were invoked for training and 1 set for testing, iterated for 100 times. The classification accuracy, precision, recall, and F-score are reported.

3. Results

3.1. Selected DTW-Based Features for All Patients. Out of the 19 calculated features, using the stepwise method, five features were found to be the ones most related to the TRS scale, containing the most amount of information related to the motor states. Three of the features were the distance measures calculated using the right foot acceleration data in the X-axis (feature 1), the Y-axis (feature 2), and using left foot acceleration data in the Z-axis (feature 3). Two additional features were the distance measures calculated using left foot gyroscope data in the X-axis (feature 4) and the magnitude of left foot gyroscope data (feature 5). The selected features show that the calculated distances from all single axes and from both acceleration and gyroscope measures were relevant to the motor states.

3.2. Evaluation of DDS

3.2.1. DDS Over the Course of Test Occasions. To investigate the extent that mean DDSs agree with the TRS scores for all PD patients, the mean DDS and the mean TRS scores across all test occasions are calculated and illustrated in Figure 4. In this figure, since DDS was calculated between every two consecutive tests, the mean DDS is visualized between the tests. The last test is excluded since during that test there was not enough data for the analysis. The large numbers in the X-axis are the time points when the tests were experimented starting from the first (baseline) test occasion, which was around 50 minutes before dose administration, denoted by -50. The follow-up test occasions were 0 minutes, the time when the dose was administered, and follow-up occasions till 320 minutes after that. The number of tests (n) was as follows: -50 (19), 0 (19), 20 (19), 40 (19), 60 (19), 80 (19), 110 (19), 140 (19), 200 (19), 230 (18), 260 (15), 290 (13), 320 (11), and 350 (1). The small numbers in the X-axis refer to the middle of the experiment time points where the mean DDS fits.
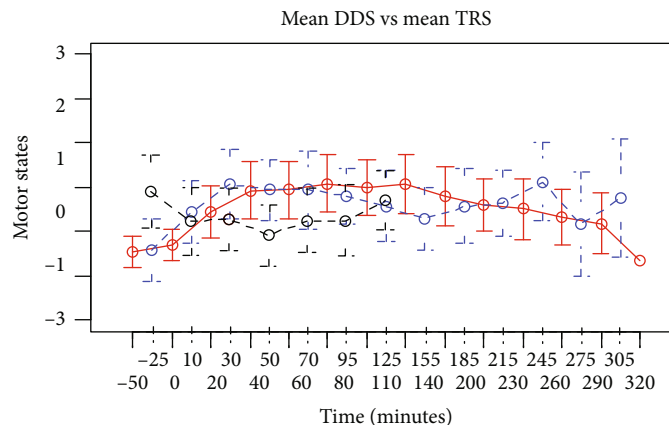
FIGURE 4: Mean DDS for PD patients (dashed, blue line) vs. mean DDS for healthy controls (dashed, black line) vs. mean TRS scores (straight, red line) across the time points ± 0.95 confidence intervals of the mean for each test occasion. None of the mean DDS for patients, mean DDS for healthy controls, and mean TRS scores was significantly different from each other.

TABLE 3: Selected DTW-based features for the most responsive patients together with their correlations and P values. In this table, acceleration data is abbreviated as Acc, gyroscope data as Gyr, right foot as R, and left foot as L. X, Y, and Z are the axes from which the measurements were recorded.

Patient ID	Feature	Correlation	P value
7	R_Magnitude_Gyroscope	0.83	≤0.001
13	L_Acceleration_Y-axis	0.55	0.07
15	R_Gyroscope_X-axis	0.75	0.01
33	L_Acceleration_Z-axis	0.65	0.12
39	L_Gyroscope_X-axis	0.63	0.05
40	Mean_RL_Acceleration_Y-axis	0.62	0.02
41	L_Gyroscope_X-axis	0.63	0.02

When depicting the mean DDS vs. mean TRS across the time points, there was an error of ±1 between DDS and TRS throughout the first to the last intervals. The overall trend of the DDS for PD patients was similar to the trend of TRS indicating DDS provides information related to motor states of PD patients. This can be specifically seen in Figure 4 between the tests -25/-50 and 140/155. In contrast to mean DDS of the PD patients and mean TRS, the mean DDS for healthy controls showed a different trend.

3.3. DTW-Based Measure for Selected Patients. Table 3 illustrates the selected patients, DTW-based features, correlation of the selected features to their TRS scores, and significance of the correlations (P value). Most of the features were from a single axis and a single foot, except two of them (magnitude of gyroscope from right foot; mean of the acceleration in the Y-axis from the right and left feet). Features showed medium to high correlations (0.55 to 0.83) with the respective TRS scores, and all were significant except two (P values = 0.12 and 0.07). The highest correlation was 0.83 where the selected feature was the magnitude of the orientation extracted from the right foot. It can also be seen in Table 3 that for these patients the left foot is selected more often than

the right foot. However, the total number of patients was not enough to draw an inference. In addition, as it was investigated in a previous study, PD asymmetry did not have an effect on the performance of the leg agility tests [10].

When investigating the most relevant DTW-based features for the most responsive PD patients according to TRS, there were different features identified.

The overlay of the scores from the selected DTW-based features (blue lines) and TRS scores (red lines) across the test occasions for the seven patients is depicted in Figure 5.

As depicted in Figure 5, the DTW-based features agree with the respective TRS scores of the responsive patients. According to this figure and the correlation presented in Table 3, the best match is shown for patient ID 7 ($R = 0.83$). The weakest correlation was between the DTW-based feature and TRS scores of patient ID 13 ($R = 0.55$). Despite the low correlation, the DTW-based and TRS scores for this patient had a similar trend. On the other hand, patient ID 33 had a good correlation ($R = 0.65$), but it was not significant since in clinical experts' rating the TRS response was not manifested right after the medication administration at test 2, whereas the DTW-based feature shows there was some response at earlier intervals (test 3). This feature (Acc_Z-axis) might have captured the motor state change in micromovements along the Z-axis that was not visible to clinical experts' eyes. For patient ID 41 and ID 15, even though mean TRS responses were not large, the DTW-based scores matched well with the respective TRS scores. Extraction of DTW-based features using acceleration or gyroscope signals from individual axes showed promise for measuring the motor states in PD. The results indicate that this method is able to provide motor state information at an individual level. However, the number of responsive PD patients in this study was limited and DDS needs to be calculated for a larger number of such cases.

3.4. Analysis of DDS for Asymmetric PD Patients. All PD patients had asymmetrical motor symptoms; nine PD patients were affected mostly on the right side and 10 PD patients on the left. In the two groups of PD patients, the first PC between right and left legs was not significantly different.
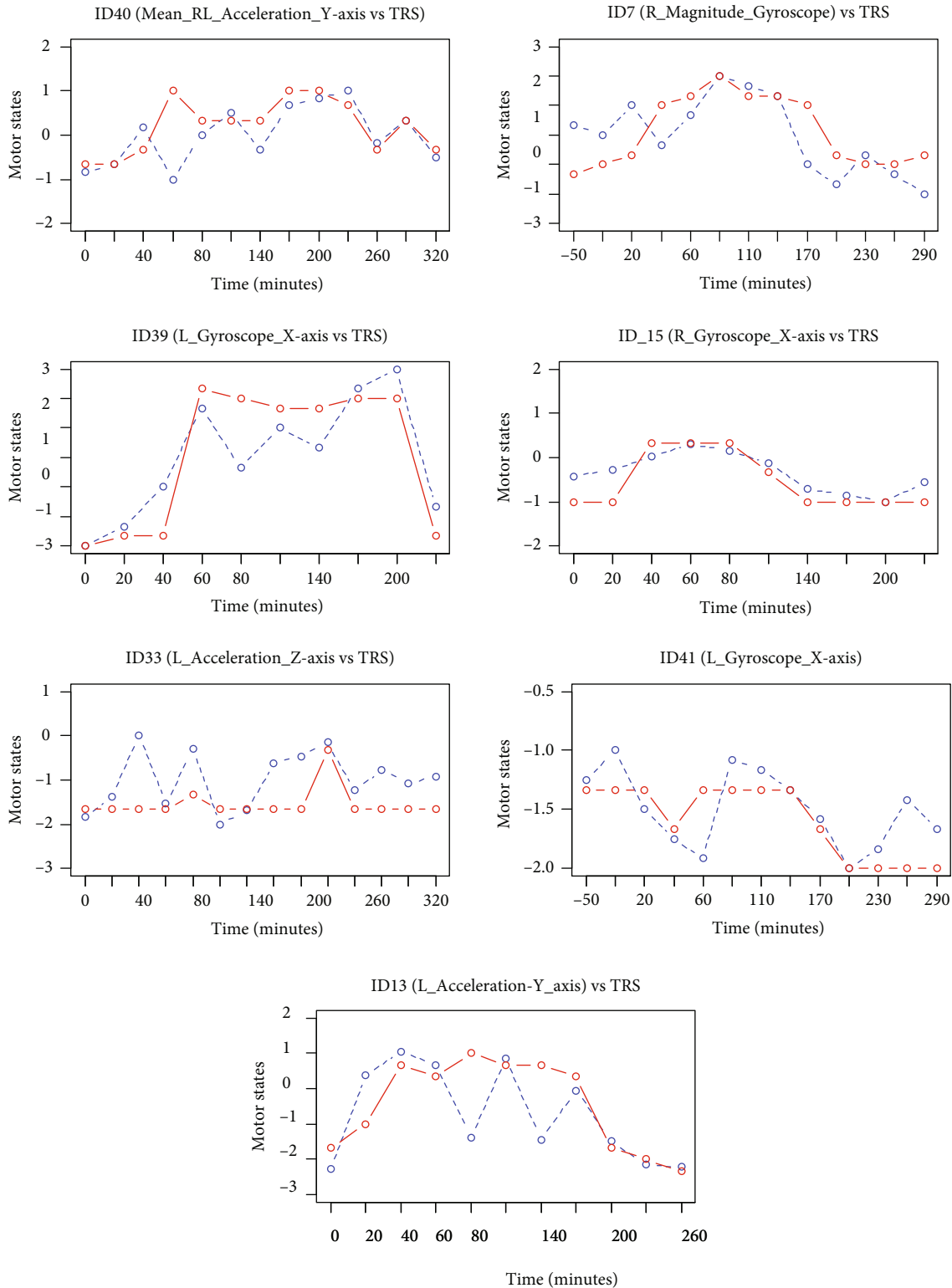
FIGURE 5: The calculated and selected DTW-based measures (blue) vs. the TRS scores (red), for seven responsive patients, depicted across the test occasions.

P value was 1 for both groups of right and left affected PD patients. This indicates asymmetry of PD did not reflect on performing the leg agility tests nor was it detected by DDS, as measured by the first PC.

3.5. Mean DDS vs. TRS Groups and Healthy Controls. The power of DDS to differentiate between groups of PD patients at different motor states was assessed. Figure 6 shows the result of assessing the differences in mean DDS between tests
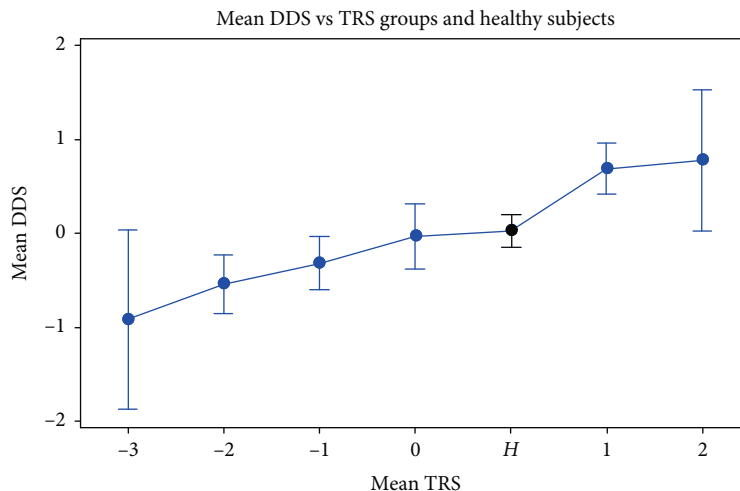
FIGURE 6: Comparison of means for DDS across groups of patients at different motor states scored by clinical TRS. In the Y-axis, the negative values mean Off state and the positive values mean On state. H is the group of healthy subjects. Number of observations (n): -3 (5), -2 (47), -1 (57), 0 (38), H (153), 1 (62), 2 (8), and 3 (0). The DDSs were significantly different between 1 and -3 ($P = 0.02$), 1 and -2 ($P \leq 0.001$), 1 and -1 ($P \leq 0.001$), 1 and 0 ($P = 0.02$), 1 and H ($P \leq 0.001$), 2 and -2 ($P = 0.02$), and H and -2 ($P = 0.03$).

TABLE 4: Classification accuracy, precision, recall, and F-score results for other rating scales.

Rating scale items	SVM				Decision Trees			
	Accuracy	Precision	Recall	F-score	Accuracy	Precision	Recall	F-score
TRS	82%	0.67	0.82	0.74	74%	0.54	0.73	0.62
Dyskinesia	87%	0.75	0.87	0.80	86%	0.75	0.86	0.80
Bradykinesia (#31)	88%	0.77	0.88	0.82	87%	0.77	0.87	0.82
Gait (#29)	88%	0.77	0.88	0.82	87%	0.77	0.88	0.82
Arising from chair (#27)	91%	0.83	0.91	0.86	91%	0.83	0.91	0.86
Leg agility (#26)	83%	0.68	0.82	0.75	82%	0.68	0.82	0.75

experimented by patients at different motor states. The mean of the DDS value from left to right for each group was -0.91, -0.54, -0.32, -0.03, 0.02, 0.69, and 0.78. The mean DDS was significantly different between the patients with TRS score of 1 (mild dyskinesia) and the patients with TRS scores of 0, -1, -2, and -3 with respective P values of 0.02, $P \leq 0.001$, $P \leq 0.001$, and 0.02. Moreover, the mean DDSs were significantly different between groups of patients with a TRS score of 2 and -2 (P value = 0.02). This indicates that DDS can separate the PD patients at "Off" state from the ones at "On" state. In addition, as depicted in Figure 5, the mean DDS for healthy controls (0.02) is similar to the mean DDS for group of patients with a TRS score of 0 (-0.03). This indicates the ability of this method in identifying the On/normal mobility as it was shown by healthy subjects, experienced by patients at On state and scored by clinicians.

3.6. *Classification of Motor State Changes.* DDS was able to classify the motor state change as scored by TRS with an accuracy of 82% and 74% using the SVM and DT algorithms, respectively. The details of the classification results such as precision, recall, and F-score are presented in Table 4. The classification power of DDS was further examined for the calculated changes in other UPDRS III items.

The results indicate that DDS is able to classify the motor state changes as assessed by TRS and UPDRS III items.

4. Discussion

In this study, four main analyses were done to investigate the feasibility and the extent the DTW method is able to provide information about motor states in advanced PD. (i) The trend of DDS was examined against the TRS scores over the course of the test occasions from before medication up to 15 times after the medication; (ii) individual DTW-based features matching the TRS scores of the individual PD patients were identified; (iii) the mean DDS was compared between groups of PD patients with different motor states as they were scored by clinical experts, including the healthy subjects; (iv) the classification power of the DDS in detecting the changes in motor states was examined.

The novelty of the approach proposed in this study was that we calculated the distances between every two consecutive tests experimented at Off, On/dyskinetic, and wearing off motor states by applying the DTW method on raw signals of leg agility experiments from PD patients. The relative distance measures from the first test provided information about how the motor states of the patients changed over time.

The trends of the DDS for PD patients and the TRS scores over the test occasions were matching well up to the sixth test after taking the medication, whereas the mean DDS of the healthy subjects differentiated them by a symmetrically different trend. When investigating the best match among the 19 individual DTW-based features calculated for each patient, we found that for the patients who showed a response to TRS, there is a feature that matches well with their TRS scores. The dissimilarity of the features can refer to the multidimensionality of the PD where symptoms manifest at different limbs, different extents, and directions. This also indicates that PD patients require personalized motor state assessment that needs to be done by using machine learning algorithms when a full set of data-driven and knowledge-driven measures are provided to them. Investigating DTW further for detecting the changes in motor states as scored by TRS provided good accuracy. The results from analysis of PD asymmetry showed that it did not affect the leg agility performance as it was measured by DTW. This result is in line with our previous study where in the two groups of PD patients the first PC was not significantly different between each leg. The differences in TRS scoring of the patients in this study were not large. The mean DDS differentiated the patients at On/dyskinesia and the patients at Off states well (Figure 6). However, the number of patients during the states with larger TRS scores of two and three was not enough. In addition, the mean DDS for healthy controls matched well with the mean DDS of the PD patients at TRS score of zero, indicating the power of this method in matching the normal mobility of PD patients and healthy controls. The classification power of DTW in detecting the changes in motor states was good. However, not all PD patients in this study showed response to medication; hence, the state changes were not large as well. This should be examined including many responsive PD patients according to their TRS scores.

The results from application of DTW on walking and hand rotation signals were not as promising as the ones reported in this paper for leg agility. This might be because application of DTW on lengthy signals may not optimally reflect the distances providing information about motor states. To extract motor state information in this study, the distances between every consecutive signals were calculated. An alternative approach was setting the first test as baseline test when the PD patients were assumed to be at Off states and calculating the distances between every other test against the baseline. However, examination of this approach was not promising. Investigating the rationale behind it, we observed that TRS score for the first tests of the PD patients was not small enough for the DTW method in order to capture the differences between this test and the latter tests. The appealing aspect of the DTW method in this study was that the analysis of data-driven distance measures extracted from consecutive tests, regardless of their length and frequency, could provide information about an important state in PD, namely, the motor states. The results of this study are in line with what was recommended by Shokoohi et al. [29]. They investigated two approaches to compute the multidimensional DTW score. In an approach named "dependent," the

distance between each corresponding pair of time series (between X, Y, and Z axes) was calculated, whereas in the "independent" approach, the score was computed for each dimension ($X/Y/Z$ axes) independently across the tests. It was recommended that with the case of right and left upper-/lower limb experiment, the choice of independent DTW score calculation would provide more accurate results for classification.

The previous study [10] provided high convergence validity (0.81) for automatic scoring of the motor states using machine learning algorithms. 24 features were extracted using various statistical methods, and ten features were selected as the most related ones to TRS. One of the selected features was approximate entropy of magnitude orientation that was calculated using a method that took the timing variability of the signals into account [30]. It could be of interest to investigate the importance of a calculated DTW-based feature along with those features in relation to TRS and to investigate whether the inclusion of this feature adds up to the convergent validity of the machine learning methods.

A limitation of this work was that extracting the distance measures requires accurately presegmented signals. This is because for calculating the distance measures using the DTW method the endpoints of the two signals must not be largely variable. This can affect the calculated measure to a large extent as it is discussed in the work of Shokoohi et al. [29], and we experimented it with walking and hand rotation data. In this study, the segmentation was done visually by cutting the signal from maximum two seconds before the start and after the end of the tasks. The agreement between the raters was 0.82 indicating that the specialist did not fully agree on scoring the motor states of the PD patients. Two clinicians rated more towards the Off state than dyskinesia which perhaps was the reason for having the range of the mean TRS scores over time towards the Off state.

For future studies, examining this method for differentiating between PD patients at different motor states requires a larger number of PD patients with a complete range of motor state ratings of -3 to +3. Moreover, investigating the method for the most responsive patients and identifying the reasons for the appearance of the individual DTW-based features for them require using a larger number of responsive PD patients to be examined using this method. Expanding this method to be used with a larger dataset might prove that it is enough to use data from only one foot to quantify the motor states. Additional sensor data such as pressure measurements could be used to further improve the method. This has also been proposed in a study by Steinmetzer et al. [31]. An alternative approach to DTW can be Hidden Markov Model (HMM) for assessing the quality of the movements as presented by Rybarczyk et al. [32]. They assessed the quality of the movements using DTW and HMM on measured body joint angles.

5. Conclusion

Using the DTW method, the calculated distance measures were able to assess the motor states similarly to visual evaluation by clinical experts using TRS. The identified DTW-

based individual features matched well with the motor states of the responsive patients. The results from evaluating the DTW-based scores at different motor states showed that this method can differentiate the patients at On state from the ones at Off states. DTW-based scores also showed high classification power when classifying the motor state changes in TRS, as well as the changes in the scoring of the other rating scales.

In conclusion, the results from this study showed it is feasible to use the DTW method for extracting information about PD motor states. The information provided by using DTW method can be included in the development of the methods for automatic scoring of advanced PD motor states.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

The patients' data used in this work were evaluated and approved by the Regional Ethic Committee in Uppsala, Sweden (protocol number 2015/100). They are in accordance with the six laws (no. 2003:460) on ethical testing of research relating to people prescribed in the application.

Disclosure

An earlier version of this study is presented as a seminar in Complex Systems & Microdata Analysis at Dalarna University.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Supplementary Materials

Figure 1: comparison of means for DDS extracted from hand rotation data across groups of patients at different motor states scored by clinical TRS. In the Y-axis, the negative values mean Off state and the positive values mean On state. H is the group of healthy subjects. DDS was not significantly different between the groups. Figure 2: comparison of means for DDS extracted from walking data across groups of patients at different motor states scored by clinical TRS. In the Y-axis, the negative values mean Off state and the positive values mean On state. H is the group of healthy subjects. DDS was not significantly different between the groups. Figure 3: DDS scores were calculated between signals of each test vs. first test (baseline) using the leg agility data. Mean DDS for PD patients (dashed, blue line) vs. mean TRS scores (straight, red line) across the time points ± 0.95 confidence intervals of the mean for each test occasion. None of the mean DDS for patients, mean DDS for healthy controls, and mean TRS scores was significantly different from each other. Mean DDS did not show a similar trend to mean TRS scores. *(Supplementary Materials)*

References

[1] G. DeMaagd and A. Philip, "Parkinson's disease and its management: part 1: disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis," *P & T: A Peer-Reviewed Journal for Formulary Management*, vol. 40, no. 8, pp. 504–532, 2015.

[2] F. M. Navarro-Peternella and S. S. Marcon, "Quality of life of a person with Parkinson's disease and the relationship between the time of evolution and the severity of the disease," *Revista Latino-Americana de Enfermagem*, vol. 20, no. 2, pp. 384–391, 2012.

[3] J. G. Nutt, S. T. Gancher, and W. R. Woodward, "Does an inhibitory-action of levodopa contribute to motor fluctuations," *Neurology*, vol. 38, no. 10, pp. 1553–1557, 1988.

[4] W. Maetzler, J. Klucken, and M. Horne, "A clinical view on the development of technology-based tools in managing Parkinson's disease," *Movement Disorders*, vol. 31, no. 9, pp. 1263–1271, 2016.

[5] M. Merello and A. J. Lees, "Beginning-of-dose motor deterioration following the acute administration of levodopa and apomorphine in Parkinson's disease," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 55, no. 11, pp. 1024–1026, 1992.

[6] C. G. Goetz, S. Fahn, P. Martinez-Martin et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): process, format, and clinimetric testing plan," *Movement Disorders*, vol. 22, no. 1, pp. 41–47, 2007.

[7] D. Nyholm, A. I. M. Nilsson Remahl, N. Dizdar et al., "Duodenal levodopa infusion monotherapy vs oral polypharmacy in advanced Parkinson disease," *Neurology*, vol. 64, no. 2, pp. 216–223, 2005.

[8] D. Isacson, K. Bingefors, I. S. Kristiansen, and D. Nyholm, "Fluctuating functions related to quality of life in advanced Parkinson disease: effects of duodenal levodopa infusion," *Acta Neurologica Scandinavica*, vol. 118, no. 6, pp. 379–386, 2008.

[9] T. T. Um, F. M. J. Pfister, D. C. Pichler et al., "Parkinson's disease assessment from a wrist-worn wearable sensor in free-living conditions: deep ensemble learning and visualization," 2018.

[10] S. Aghanavesi, F. Bergquist, D. Nyholm, M. Senek, and M. Memedi, "Motion sensor-based assessment of Parkinson's disease motor symptoms during leg agility tests: results from levodopa challenge," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 111–119, 2020.

[11] M. D. Hssayeni, M. A. Burack, and B. Ghoraani, "Automatic assessment of medication states of patients with Parkinson's disease using wearable sensors," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6082–6085, Orlando, FL, USA, August 2016.

[12] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[13] J. B. Donald and C. James, "Using dynamic time warping to find patterns in time series," in *Association for the Advancement of Artificial Intelligence, Workshop on Knowle Discovery in Databases (AAAI)*, pp. 229–248, Seattle, WA, USA, 1994.

[14] T. Fu, F. Chung, R. Luk, and C. M. Ng, "Stock time series pattern matching: template-based vs. rule-based approaches," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 3, pp. 347–364, 2007.

[15] N. H. Ghassemi, J. Hannink, C. Martindale et al., "Segmentation of gait sequences in sensor-based movement analysis: a comparison of methods in Parkinson's disease," *Sensors*, vol. 18, no. 2, p. 145, 2018.

[16] J. Barth, C. Oberndorfer, P. Kugler et al., "Subsequence dynamic time warping as a method for robust step segmentation using gyroscope signals of daily life activities," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6744–6747, Osaka, Japan, July 2013.

[17] J. G. Barth, *Development and Validation of a Mobile Gait Analysis System Providing Clinically Relevant Target Parameters in Parkinson's Disease*, FAU University Press, Erlangen, 2017.

[18] Vikas and R. K. Sharma, "Analysis of voice for Parkinson's disease persons using dynamic time warping technique," in *Intelligent Computing, Communication and Devices*, vol. 308 of Advances in Intelligent Systems and Computing book series, pp. 753–759, Springer India, New Delhi, 2015.

[19] M. R. Adame, A. Al-Jawad, M. Romanovas et al., "TUG test instrumentation for Parkinson's disease patients using inertial sensors and dynamic time warping," *Biomedical Engineering / Biomedizinische Technik*, vol. 57, pp. 1071–1074, 2012.

[20] M. Giuberti, G. Ferrari, L. Contin et al., "On the characterization of leg agility in patients with Parkinson's disease," in *2013 IEEE International Conference on Body Sensor Networks*, pp. 1–6, Cambridge, MA, USA, May 2013.

[21] F. Parisi, G. Ferrari, M. Giuberti et al., "Body-sensor-network-based kinematic characterization and comparative outlook of UPDRS scoring in leg agility, sit-to-stand, and gait tasks in Parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1777–1793, 2015.

[22] M. Senek, S. M. Aquilonius, H. Askmark et al., "Levodopa/carbidopa microtablets in Parkinson's disease: a study of pharmacokinetics and blinded motor assessment," *European Journal of Clinical Pharmacology*, vol. 73, no. 5, pp. 563–571, 2017.

[23] C. A. Ratanamahatana and E. J. Keogh, "Making time-series classification more accurate using learned constraints," in *Proceedings of the 2004 SIAM International Conference on Data Mining*, Lake Buena Vista, FL, USA, April 2004.

[24] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[25] F. Javed, I. Thomas, and M. Memedi, "A comparison of feature selection methods when using motion sensors data: a case study in Parkinson's disease," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5426–5429, Honolulu, HI, USA, July 2018.

[26] M. Ringner, "What is principal component analysis?," *Nature Biotechnology*, vol. 26, no. 3, pp. 303-304, 2008.

[27] I. Steinwart and A. Christmann, *Support Vector Machines*, Springer Publishing Company, Incorporated, 2008.

[28] R. Kohavi and J. R. Quinlan, *Data Mining Tasks and Methods: Classification: Decision-Tree Discovery*, Oxford University Press, Inc., 2002.

[29] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data Mining and Knowledge Discovery*, vol. 31, no. 1, pp. 1–31, 2017.

[30] S. Aghanavesi, M. Memedi, M. Dougherty, D. Nyholm, and J. Westin, "Verification of a method for measuring Parkinson's disease related temporal irregularity in spiral drawings," *Sensors*, vol. 17, no. 10, article 2341, 2017.

[31] T. Steinmetzer, I. Bonninger, B. Priwitzer et al., "Clustering of human gait with Parkinson's disease by using dynamic time warping," in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, pp. 1–6, San Carlos, Costa Rica, July 2018.

[32] Y. Rybarczyk, J. Luis Pérez Medina, L. Leconte, K. Jimenes, M. González, and D. Esparza, "Implementation and assessment of an intelligent motor tele-rehabilitation platform," *Electronics*, vol. 8, no. 1, p. 58, 2019.