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

Federico Parisi, *Student Member, IEEE*, Gianluigi Ferrari, *Senior Member, IEEE*, Matteo Giuberti, *Member, IEEE*, Laura Contin, Veronica Cimolin, Corrado Azzaro, Giovanni Albani, and Alessandro Mauro

Manuscript received April 15, 2015; revised July 15, 2015; accepted August 16, 2015. Date of publication August 25, 2015; date of current version November 3, 2015. This work was supported in part by the Italian Ministry of Health (RF-2009-1472190).

F. Parisi and G. Ferrari are with the CNIT Research Unit of Parma and the Department of Information Engineering, University of Parma, I-43124 Parma, Italy (e-mail: federico.parisi@studenti.unipr.it; gianluigi.ferrari@unipr.it).

M. Giuberti was with the CNIT Research Unit of Parma and the Department of Information Engineering, University of Parma, I-43124 Parma, Italy. He is now with Xsens Technologies B.V., 7500 AN Enshede, The Netherlands (e-mail: matteo.giuberti@xsense.com).

L. Contin is with Research & Prototyping, Telecom Italia, 10148 Turin, Italy (e-mail: laura.contin@telecomitalia.it).

V. Cimolin is with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milano, Italy (e-mail: veronica.cimolin@ polimi.it).

C. Azzaro and G. Albani are with the Division of Neurology and Neurorehabilitation, Istituto Auxologico Italiano, IRCCS, I-28824 Piancavallo (VB), Italy (e-mail: c.azzaro@auxologico.it; g.albani@auxologico.it).

A. Mauro is with the Division of Neurology and Neurorehabilitation, Istituto Auxologico Italiano, IRCCS, I-28824 Piancavallo (VB), Italy, and also with the Department of Neurosciences, University of Turin, 10125 Turin, Italy (e-mail: mauro@auxologico.it).

Color versions of one or more of the figures in this paper are available online.

I. Introduction

PARKINSON'S disease (PD) is a progressive, chronic, neurodegenerative condition that is responsible for a gradual motor impairment. Therapies based on the use of dopaminergic drugs, such as L-dopa, are useful to manage the early PD motor symptoms, but their efficacy worsens over time, causing additional motor complications, such as dyskinesia and motor fluctuations, which can further impair the patients' life quality. An accurate and continuous monitoring of the symptoms' progression and treatment effect should be required in order to define an effective therapy, but neurologists can often rely only on qualitative and sporadic clinical observations, which may not be representative of the actual disease status. A more objective evaluation can be achieved using semiquantitative rating scales, such as the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) [1]. Although several studies have pointed out a good test–retest reliability for the global UPDRS motor score, the latter has some limits in clinical trials. Some of these limitations are: the need for a trained neurologist to assign the UPDRS score; the interrater/intrarater variability [2]; the discrete nature of the UPDRS (scores from 0 to 4), which is not optimal to detect minimal changes during the disease progression [3]; and the difficulty to convey a concise score, especially when several movement components (such as speed, amplitude, hesitations, etc.) should be taken into account for the evaluation. In order to improve the UPDRS assessment reliability, raising the detection rate between disease-modifying and symptomatic effects over a specific treatment regime, the sample size and the clinical trial duration should be increased [4]. Obviously, this solution is not always feasible or practical. The clinical judgment of the disease stage in PD patients has benefited from having the patients keep motor diaries while at home. However, this tool is often unreliable because of nonoptimal compliance in the patient record keeping, recall bias, or weak self-assessment skill due to cognitive impairment, which is often associated with late stages of PD [5].

The advent of new motion-sensing devices, which allow accurate movement measurement through a kinematic analysis, has enabled the design of "man versus machine" clinical trials, with relevant implications in practical terms. In recent years, the use of integrated computerized systems, such as body sensor networks (BSNs), has proliferated in clinical environments, allowing easier acquisition of objective and quantitative measurements that can be repeated several times on a daily basis at the discretion of both patients and neurologists [6].

Recently, a unified approach to the evaluation of specific UPDRS motor tasks, namely, the leg agility (LA) [7], the sit-to-stand (S2S) [8], and gait (G) tasks [9], [10], has been proposed. The designed system relies on a simple BSN formed by three inertial measurement units (IMUs) (two on the thighs and one on the chest) and aims at characterizing the considered tasks by extracting and analyzing the kinematic features associated with their typical movement patterns, in both time and frequency domains. The automatic data acquisition sessions have been carried out concurrently with clinical evaluations, performed by neurologists with expertise in PD and according to the standards of the MDS. The extracted features and the subjective evaluations of the neurologists were then used to train an automatic UPDRS scoring system, with the aim to automatically assess the patients' motor performance matching as closely as possible the medical evaluation criteria.

Unlike our previous studies [7]–[10] and the majority of the existing literature, in which UPDRS tasks are analyzed singularly, in this paper, we focus on the comparative evaluation of the LA, S2S, and G tasks. An experimental analysis of the data from 34 PD patients and the UPDRS evaluations of three expert neurologists was carried out. The most relevant features, for the kinematic characterization of each task, were identified and motor performance of patients belonging to the different UPDRS classes was analyzed. We also considered the performance achieved by the designed automatic classification system in the three tasks, proposing a comparative outlook with the interneurologist assessment. Then, we investigated the correlation between the UPDRS scores assigned to the tasks by both the neurologists and our automatic system, highlighting that the latter shows a performance which is compliant with typical interneurologist variability, i.e., it behaves correctly. Furthermore, we introduce an aggregate UPDRS score, simply defined as the sum of the scores obtained in each task, as a significant concise metric which can provide additional information to neurologists for deriving insights on the overall level of impairments of patients and on the relative "weight" of each task in the assessment of the gravity of the symptoms. Finally, the feasibility of an application for remote rehabilitation and monitoring of PD patients in a telemedicine environment is discussed, and a possible efficient implementation approach is proposed.

The structure of the paper is the following. In Section II, preliminaries and an overview on related works are given. The hardware configuration and the set of subjects considered in the experiments are presented in Section III. In Section IV, we describe the methods used for the kinematic features extraction through the inertial BSN in each task and for the automatic UPDRS scoring system implementation. The experimental results are shown in Section V and discussed in Section VI, together with a possible application of the proposed system for the management of PD patients in a telemedicine scenario. Conclusions are presented in Section VII.

II. Preliminaries and Related Work

A. UPDRS Tasks

The guidelines of the MDS for the evaluation of PD motor tasks are described in the Part III of the UPDRS document [1]. In this paper, we focus on the items 3.8, 3.9, and 3.10, which correspond to the LA, Arising from Chair,[1] and G tasks. The choice of these particular tasks was influenced by the need to keep the BSN as simple as possible, maximizing at the same time the number of tasks which could be analyzed without changing the sensors' placement. The selected tasks are particularly suitable for the considered unified analysis and clinically relevant for a comprehensive evaluation of the patients' symptoms, as they refer to different aspects of PD (for example, LA is related to bradykinesia, while S2S and G are associated with posture/deambulation symptoms). We now quickly summarize the basic characteristics of each of these tasks.

1) LA Task: The LA exercise consists of alternately raising up and stomping the feet on the ground, as high and as fast as possible. Ten repetitions per leg must be performed while sitting on the chair in order to test each leg separately (in the following, we will distinguish between right LA (RLA) and left LA (LLA) tasks).[2] The significant parameters that have to be measured, independently for each leg, are the speed, the regularity, and the amplitude of the movement.

2) S2S Task: In the S2S task, the patient is asked to sit on a straight-backed chair with armrests. The exercise consists in crossing the arms across the chest (in order to avoid their use in the movement) and getting up from the chair. In the case of failure, the patient can retry to raise up to two more times. If still unsuccessful, the patient can move forward on the chair to facilitate the movement, or in case of another failure, he/she can use the armrests to stand up.

3) G Task: In the G task, the patient is asked to walk, at his/her preferred speed, away from the examiner for at least 10 m and in straight line, then to turn around and return to the starting point. The parameters of interest are those strictly related to the gait characteristics, such as the stride/step amplitude and speed, the cadence, the gait cycle time (GCT), parameters related to the turning phase, the variability between left and right steps, and the arm swing. Freezing of gait should be evaluated separately. The assessment of the upper limbs (e.g., the arm swing) will not be considered in this paper, as no sensor is placed on the arms in the designed BSN. An extension of the current approach including sensing devices on the arms represents an interesting research direction.

4) UPDRS Evaluation: As mentioned in Section I, the UPDRS allows assignment of an integer score to a patient's motor

[1]For consistency with our previous work [8], in the following, we denote the Arising from Chair task as S2S task.

[2]When not specified, LA refers to the general task, including both RLA and LLA trials.

TABLE I
UPDRS MAPPING FOR THE CONSIDERED TASKS

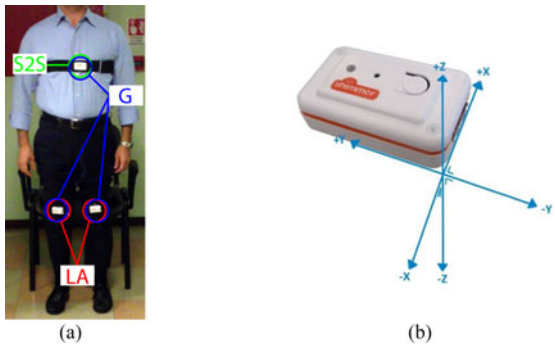
LA task	UPDRS	Amplitude	Hesitations	Interruptions	Slowing	Freezing
	0	nearly constant	no	0	0	0
	1	decrements near the end	slight	≥ 1	1,2	0
	2	decrements midway	mild	–	3,4,5	0
	3	decrements after first tap	moderate	–	≥ 6	≥ 1
	4	always minimal or null	severe	–	always	–
S2S task	**UPDRS**	**Failed attempts**		**Use of armrests**	**Slowing**	**Move forward on chair**
	0	0 failed attempts		no	no	no
	1	≥ 1 failed attempts		no	yes	yes
	2	0 failed attempts		yes	–	–
	3	≥ 1 failed attempts		yes	–	–
	4			not able to stand up alone		
G task	**UPDRS**	**Independent walking**		**Impairments level**		
	0	yes		no impairments		
	1	yes		minor impairments		
	2	yes		substantial impairments		
	3	no		assistance device needed for safe walking		
	4	no		cannot walk at all or only with another person's assistance		

performance in a specific task, ranging from 0, which means that the patient is able to perform the task normally and with no impairments, to 4, which means that the patient has major difficulties in performing the exercise or is not able to perform it at all. Table I maps the characteristics of each task which the examiner should consider for the assessment of the patient's performance to the UPDRS scores.

B. Related Work

The kinematic analysis of specific motor tasks through different motion capture and sensing technologies, such as optoelectronic systems and inertial-based BSN, has been widely studied for various clinical applications. With regard to PD, the majority of the existing literature has focused on the quantitative kinematic characterization of single motor tasks and/or on the evaluation of patients' performance in different PD conditions. In this context, the LA task [7], [11], the S2S task [8], [12], the G task [10], [13], [14], and tremors [15] have been analyzed.

To the best of our knowledge, only limited attention has been devoted to investigation of the relationship between different UPDRS tasks characterized through motion capture technologies. Stochl *et al.* [16] investigated the structure of PD symptoms in terms of the motor symptom evaluations defined in UPDRS Part III. Five main latent symptoms factors were identified, and the correlations between the UPDRS scores assigned to the various tasks were reported. Similarly, in [17], a statistical analysis of the UPDRS motor scores was performed, using classical evaluation methods by neurologists and considering PD patients in both ON (i.e., the intervals during which the medication is effective) and OFF (i.e., the intervals during which the medication is not effective) conditions, in order to identify latent relationships between UPDRS tasks and combine the tasks in "macrogroups" related to similar PD symptoms. Five of these groups, denoted as "factors," have been identified (namely, gait/posture, tremor, rigidity, left extremities bradykinesia, and right extremities bradykinesia). The correlations 1) between the UPDRS scores assigned by neurologists to the introduced "macro-groups" of tasks and 2) between them and an aggregate UPDRS score are also presented, showing that the macrogroups can be assessed separately and provide information about different aspects of the disease. From our point of view, carrying out a comparative analysis of the LA, S2S, and G tasks, using 1) the same BSN for the kinematic characterization of each task and 2) the same approach for the design of an automatic UPDRS scoring system, allows analysis of the correlations between the UPDRS values of single tasks and a total UPDRS score, thus opening an interesting research direction.



(a) (b)

Fig. 1. (a) Inertial BSN designed for the evaluation of the three UPDRS tasks of interest (LA, S2S, G): the subsets of nodes used in each task are marked with different colors. (b) Shimmer device (IMU) and its reference coordinate system.

III. EXPERIMENTAL SETUP

A. Hardware

The BSN designed for the unified evaluation of the LA, S2S, and G tasks is formed by only three IMUs, one on the chest and one per thigh, attached to the body with Velcro straps, as shown in Fig. 1(a). Each node is a Shimmer device (http://www.shimmersensing.com/, [18]), which is a small (dimensions: 53 mm $\times$ 32 mm $\times$ 25 mm; weight: 22 g) and low-power wireless IMU, equipped with a triaxial accelerometer, a triaxial gyroscope, and a triaxial magnetometer. A Shimmer node and its reference coordinate system are shown in Fig. 1(b). The sampling rate is set to 102.4 Hz, which is the closest, in the

set of the sampling rates supported by the Shimmer platform, to the sampling frequency (namely, 100 Hz) of the optoelectronic reference system (Vicon, Oxford, U.K.) used for the validation of the inertial signals' accuracy. The acquired data are streamed wirelessly (via a Bluetooth radio interface) to a personal computer, where signal processing and automatic classification (described in details in Section IV) are performed.

The placement of the sensors has been chosen taking into account two main motivations: 1) the need to analyze the three tasks without changing the configuration of the nodes in order to minimize the patients' stress and simplify the acquisition procedure, allowing sequential execution of the three tasks; and 2) the higher accuracy and reliability of IMUs in measuring inclinations and accelerations, rather than positions or displacements. With the current BSN configuration, indeed, all the kinematic parameters in the LA and S2S tasks are extracted from inclination and/or angular velocities measured with the nodes on the thighs and on the chest, respectively; at the opposite, in the G task, the majority of the features are extracted from acceleration signals directly measured by the sensor placed on the trunk.

For validation purposes, the data acquired with the inertial BSN and the extracted kinematic features have been compared with those measured with the Vicon optoelectronic system. In particular, in [7], we have first demonstrated the equivalence between heel' and thigh' kinematics. More precisely, the 3-D orientations of the Shimmer nodes placed on the thighs are estimated, with reference to the Earth frame, through an orientation estimation filter [19]. For each leg, the orientation component in the sagittal plane, corresponding to the inclination θ (dimension: [deg]) of the thigh, is extracted, together with the thigh's angular velocity ω (dimension: [deg/s]). These signals are then compared to those estimated with the optoelectronic system taking into account the 3-D positions and velocities of reflective markers placed on the subject's heels. The results presented in [7] show a strong correlation (approximately equal to 0.98) between heels' optical data and thighs' inertial data, motivating the use of θ and ω for the kinematic characterization of the LA task. The same approach has been applied, in the S2S task, to determine the accuracy, with respect to the Vicon system, of the trunk inclination estimated with a Shimmer node and similar accuracy results have been obtained. The validation of kinematic parameters for the G task is discussed in [9] and [10]. The average errors are comparable with those obtained in other studies in the literature, such as [13], [20], [21], and are sufficiently low to be considered almost negligible for the purpose of this study—for example, the average errors for the estimations of some temporal parameters, such as the heel-strike (HS), toe-off (TO), and GCT, are (mean $\pm$ standard deviation) 8.22 ± 17.6, 6.83 ± 26.3, and 8.87 ± 23.7 ms, respectively, whereas for spatial parameters, such as the stride length and the step length, the average errors are 4.23 ± 4.94 and 3.15 ± 7.34 cm, respectively.

B. Subjects and Acquisition Procedure

The subjects in these studies were 34 PD patients, including 22 males and 12 females with average age equal to 67.4 years (ages between 31 and 79 years) and standard deviation equal to 11.6 years. The average Modified Hoehn and Yahr Scale score for the subjects was 1.6 (standard deviation equal to 0.47, minimum score equal to 1, maximum score equal to 3) on the 1-to-5 scale (higher scores indicate more severe impairments and more advanced stages of the disease). The sensing devices were placed on a patient's body as shown in Fig. 1(a), trying to align the x-, y-, and z-axes of the node coordinate reference system, shown in Fig. 1(b), to the upward–downward, right–left, and forward–backward directions, respectively. The alignment of the sensors, with respect to the anatomical structure, is aided by the developed acquisition software. Before the beginning of the acquisition procedure, a check is performed on the sensors' placement, considering both the gravity direction and the 3-D orientation of the Shimmer nodes in the Earth frame: if the alignment is within a confidence range (heuristically defined), the examiner is allowed to proceed in the acquisition procedure; otherwise, a warning message is shown and the procedure is stopped until the sensors' placement is correctly modified by the examiner.

The data acquisitions were carried out by asking the patient to execute the LA, S2S, and the G tasks sequentially. In each task, only the signals recorded by a proper subset of nodes of the BSN were considered—the subsets of Shimmer devices used for the evaluation of the single tasks are shown in Fig. 1(a) using different colors. For the LA task acquisitions, only the two devices placed on the thighs were used, whereas for the S2S task, only the trunk-mounted node was considered. In the G task, all the BSN IMUs were used to achieve a complete characterization of the complex gait movement. Although we studied 34 patients, a total of 47 trials (94 for the LA task, considering separately RLA and LLA) per task have been acquired, since some of the patients performed the tasks in distinct PD conditions, i.e., in ON/OFF states or at different times corresponding to different motor fluctuation phases. The motor performance of the same subject in these situations and, consequently, the recorded kinematic patterns and the assessment by the neurologists may vary consistently. To avoid distortion in the results, we include in our dataset only the trials by the same patient in which substantial variations in motor performance have been observed, allowing us to consider each trial as a single sample for the following analysis. We remark that a detailed analysis of PD patients' motor performance, distinguishing between ON and OFF conditions, represents a relevant extension of our work. All the trials have been assessed independently by three neurologists expert in movement disorders, using a noninteger scale with intermediate scores ($\cdot.5$) to label the trials in which the neurologists were undecided between consecutive (integer) UPDRS classes. The consensus score, denoted as $\mathrm{UPDRS}_{\mathrm{Mean}}$, is defined as the arithmetic average of the scores assigned by the three neurologists to each trial, rounded to the nearest (integer or intermediate) UPDRS value. This methodology has been already used in the literature to combine the assessments of multiple neurologists in a single concise score [22], enhancing the robustness of the evaluation and reducing the distortion caused by the interrater variability [2]. In Fig. 2, the distributions of $\mathrm{UPDRS}_{\mathrm{Mean}}$ and of the UPDRS scores assigned by the three neurologists to the 47 trials are shown. It can be observed that there are slight

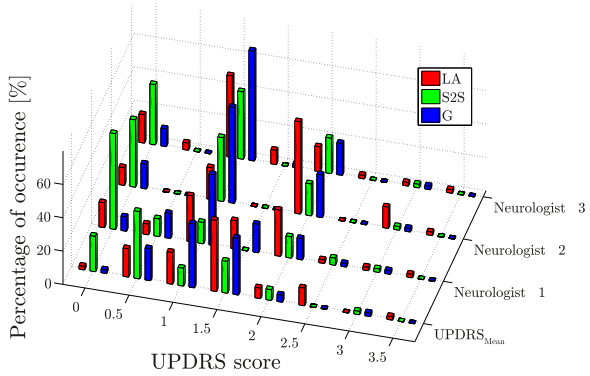


Fig. 2. Trials distribution for all the considered tasks.

differences among neurologists, in both the assessment criteria (for example the neurologist 1 uses more than the other two the intermediate scores, whereas neurologist 2 tends to assign higher scores than the other two in the LA task) and the distribution of the UPDRS scores in the three tasks. In particular, the latter seems to be a Gaussian-like distribution centered in correspondence to a dominant UPDRS class, which depends on both the neurologist and the task. Finally, the distribution of the $\text{UPDRS}_{\text{Mean}}$ score, as expected, is smoother than the UPDRS scores of the single neurologists and the scores in the LA, S2S, and G tasks show Gaussian-like distributions centered in 1.5, 0.5, and 1, respectively. We note that in an "ideal" scenario, all UPDRS classes should have similar number of patients, i.e., the UPDRS distribution should be almost uniform.

IV. Methods

A. Kinematic Features Extraction

The unified approach proposed in this study, which aims, through a single BSN, to automatically assign UPDRS scores to the LA, S2S, and G tasks, considers the same data processing and automatic classification methods presented in the single tasks analyses proposed in [7]–[10]. In the following, a concise description about the features' extraction procedure is provided. For ease of visualization, in Table II, a summary of the most relevant features identified for each task is provided. The names of the parameters considered in the experimental results are highlighted in bold.

1) LA Task: As anticipated in Section III-A, both the inclination (θ, dimension: [deg]) and the angular velocity (ω, dimension: [deg/s]) of the thighs in the sagittal plane, measured by the Shimmer nodes on the thighs, are considered for the kinematic analysis of the LA task. An illustrative portion of the inclination signal θ of one thigh, for two consecutive LA repetitions, denoted as rth and $(r+1)$th ($r \in \{1, 2, \ldots, 9\}$), is shown in Fig. 3(a). Through an automatic segmentation procedure, three fundamental time epochs, denoted as $t_{\text{S}}(r)$, $t_{\text{E}}(r)$, and $t_{\text{P}}(r)$ and associated, respectively, with the start, the end, and the epoch of maximal thigh inclination of the rth LA repetition, are identified. Starting from these epochs, various parameters can be straightforwardly calculated. The results obtained in [7] have shown that, in the time domain, the most relevant features for the kinematic characterization of the LA task are: the angular amplitude Θ (dimension: [deg]), the angular speed of execution Ω (dimension: [deg/s]), the pause between consecutive executions P (dimension: [s]), the regularity of execution R (dimension: [s]), and the repetition frequency F (dimension: [Hz]). In the frequency domain, upon the computation of the amplitude spectra of the discrete Fourier transforms of θ and ω, denoted as X_θ and X_ω, respectively, the powers of the inclination spectrum (P_{X_θ}) and the angular velocity spectrum (P_{X_ω}) are shown to be relevant features.

For the rth LA repetition ($r \in \{1, 2, \ldots, 10\}$), the expression of the most relevant features outlined in the previous paragraph is shown in the first part of Table II. When not specified, in the following, we refer to the average values of the features Θ, Ω, P, and R, obtained by averaging, over all consecutive repetitions, the values measured in each repetition and denoted, respectively, as Θ_{mean}, Ω_{mean}, P_{mean}, and R_{mean}.

2) S2S Task: The S2S task is the simplest one among those considered in this paper. For this reason, the characterization of the body movements during the execution of the task can be easily obtained considering only the inclination signal estimated through the chest-mounted sensor. As for the thighs' nodes in the LA task, the 3-D orientation of the Shimmer device placed on the trunk is estimated and the inclination of the torso, denoted as θ (dimension: [deg]), is measured [8]. The typical shape of θ, during the S2S task, is shown in Fig. 3(b). The following relevant time labels can be identified with a simple automatic segmentation procedure: 1) the starting epoch t_{S} of the S2S (i.e., when the chest starts bending forward); 2) the epoch of maximal bending of the chest t_{P} (placed around the middle of the S2S exercise); and 3) the ending epoch t_{E} of the S2S (i.e., when the chest returns in the vertical position). Once these time instants have been identified, the 12 features shown in the central part of Table II can be calculated. In particular, the features refer to the duration T (dimension: [s]), the angular amplitude Θ (dimension: [deg]), and the speed of execution Ω (dimension: [deg/s]) of the S2S exercise. Moreover, the difference D between the forward and the backward bending phases is computed for all the considered variables. In [8], the subset of features, among the 12 extracted, which has turned out to be the most significant for the characterization of the S2S task, includes T, T_{B}, T_{F}, D_T, Θ, and Ω.

3) G Task: The movements involved in the G task are inherently more complex than those associated with LA and S2S tasks, and for this reason, the characterization of gait through kinematic features is more challenging. In [9] and [10], an in-depth kinematic analysis of Parkinsonian gait is performed, considering features in both time and frequency domains. A novel approach for gait cycle phases segmentation, through a proper processing of the accelerometric signal of the chest-mounted inertial node, is presented. Considering the typical patterns in trunk accelerations shown in Fig. 3(c), the fundamental events, which identify a complete gait cycle and all the associated gait phases, namely, the HS (i.e., the instant at which the foot

TABLE II
SUMMARY OF THE MOST RELEVANT FEATURES CONSIDERED FOR EACH TASK

Task	Name	Definition	Dimension	r-value
LA	**Angular amplitude**	$\Theta(r) \triangleq \frac{\Theta_{\mathrm{A}}(r)+\Theta_{\mathrm{D}}(r)}{2}$	[deg]	n.s.
	Angular speed of execution	$\Omega(r) \triangleq \frac{\Theta_{\mathrm{A}}(r)+\Theta_{\mathrm{D}}(r)}{T(r)}$	[deg/s]	$-$**0.50**
	Pause of execution	$P(r) \triangleq t_{\mathrm{S}}(r+1) - t_{\mathrm{E}}(r)$	[s]	0.27
	Regularity of execution	$R(r) \triangleq t_{\mathrm{P}}(r+1) - t_{\mathrm{P}}(r)$	[s]	0.49
	Repetition frequency	$F \triangleq \frac{10}{t_{\mathrm{E}}(10)-t_{\mathrm{S}}(1)}$	[Hz]	-0.36
	Thigh inclination spectrum power	$P_{X_\theta} \triangleq \frac{1}{N}\sum_{h=0}^{N-1}(X_{\theta,h})^2$	adimensional	-0.46
	Thigh angular velocity spectrum power	$P_{X_\omega} \triangleq \frac{1}{N}\sum_{h=0}^{N-1}(X_{\omega,h})^2$	adimensional	-0.34
S2S	**Forwards bending duration**	$T_{\mathrm{F}} \triangleq t_{\mathrm{P}} - t_{\mathrm{S}}$	[s]	0.58
	Backwards bending duration	$T_{\mathrm{B}} \triangleq t_{\mathrm{E}} - t_{\mathrm{P}}$	[s]	0.56
	Total duration	$T \triangleq T_{\mathrm{F}} + T_{\mathrm{B}} = t_{\mathrm{E}} - t_{\mathrm{S}}$	[s]	**0.62**
	Forwards/backwards duration difference	$D_T \triangleq T_{\mathrm{F}} - T_{\mathrm{B}}$	[s]	0.47
	Forwards bending amplitude	$\Theta_{\mathrm{F}} \triangleq \theta(t_{\mathrm{P}}) - \theta(t_{\mathrm{S}})$	[deg]	0.35
	Backwards bending amplitude	$\Theta_{\mathrm{B}} \triangleq \theta(t_{\mathrm{P}}) - \theta(t_{\mathrm{E}})$	[deg]	0.25
	Average bending amplitude	$\Theta \triangleq \frac{\Theta_{\mathrm{F}}+\Theta_{\mathrm{B}}}{2}$	[deg]	0.33
	Forwards/backwards bending amplitude difference	$D_\Theta \triangleq \Theta_{\mathrm{F}} - \Theta_{\mathrm{B}}$	[deg]	0.20
	Forwards bending speed	$\Omega_{\mathrm{F}} \triangleq \frac{\Theta_{\mathrm{F}}}{T_{\mathrm{F}}}$	[deg/s]	-0.33
	Backwards bending speed	$\Omega_{\mathrm{B}} \triangleq \frac{\Theta_{\mathrm{B}}}{T_{\mathrm{B}}}$	[deg/s]	-0.21
	Average bending speed	$\Omega \triangleq \frac{\Theta_{\mathrm{F}}+\Theta_{\mathrm{B}}}{T} = \Omega_{\mathrm{F}}\frac{T_{\mathrm{F}}}{T} + \Omega_{\mathrm{B}}\frac{T_{\mathrm{B}}}{T}$	[deg/s]	-0.32
	Forwards/backwards bending speed difference	$D_\Omega \triangleq \Omega_{\mathrm{F}} - \Omega_{\mathrm{B}}$	[deg/s]	-0.14
Gait	**Gait Cycle Time**	$GCT_{R/L}(k) = HS_{R/L}(k+1) - HS_{R/L}(k)$	[s]	0.30
	Stance Time	$ST_{R/L}(k) \triangleq 100 \times \frac{TO_{R/L}(k)-HS_{R/L}(k)}{GCT_{R/L}(k)}$	% of GCT	n.s.
	Initial Double Support	$IDS(k) \triangleq 100 \times \frac{TO_L(k)-HS_R(k)}{GCT(k)}$	% of GCT	0.29
	Terminal Double Support	$TDS(k) \triangleq 100 \times \frac{TO_R(k)-HS_L(k)}{GCT(k)}$	% of GCT	n.s.
	Double Support	$DS(k) \triangleq IDS(k) + TDS(k)$	% of GCT	n.s.
	Limp	$Limp(k) \triangleq \lvert IDS(k) - TDS(k)\rvert$	% of GCT	0.30
	Step Length	$StepL_{R/L}(k) \triangleq K\,2\sqrt{2\ell h_{R/L}(k) - h_{R/L}(k)^2}$	% of height	-0.59 (mean)
	Stride Length	$SL(k) \triangleq StepL_R(k) + StepL_L(k)$	% of height	$-$**0.60**
	Step Velocity	$StepV_R(k) \triangleq \frac{StepL_R}{HS_L(k)-HS_R(k)}$	% of height/s	-0.59
	Thigh Range of Rotation	$Thigh\ RoR_R(k) \triangleq \max_i \theta(i) - \min_i \theta(i) \quad i \in GCT_R(k)$	[deg]	-0.49
	Cadence	$C \triangleq \frac{60f}{d_1}$	[steps/min]	n.s.
	Step Regularity	$R_{\mathrm{step}} \triangleq A_{\mathrm{unbiased}}(d_1)$	adimensional	-0.58
	Symmetry	$S \triangleq \frac{R_{\mathrm{step}}}{R_{\mathrm{stride}}}$	adimensional	n.s.
	Spectrum power for accelerations	$P_{a_{\mathrm{vert/y/z}}} \triangleq \frac{1}{N}\sum_{k=0}^{N-1}(X_{a_{\mathrm{vert/y/z},k}})^2$	adimensional	-0.38 (mean)
	Total spectrum power	$P_{\mathrm{sum}} \triangleq P_{a_{\mathrm{vert}}} + P_{a_{\mathrm{y}}} + P_{a_{\mathrm{z}}}$	adimensional	-0.43

The names of the parameters taken into account for the experimental results are marked in bold. In the last column, the correlation coefficients between the features and the neurologist-assigned UPDRS scores are shown (the best for each task is highlighted in bold). Correlation coefficients (r-values) with associated p-values (not shown here) greater than 0.05 are considered as nonsignificant (n.s.).

touches the ground) and the TO (i.e., the instant at which the foot leaves the ground), have been identified. Once all the HSs and the TOs for both legs are known, temporal parameters, such as gait cycle time (GCT, dimension: [s]), Stance Time (ST, dimension: [% of GCT]), double support time (DS, dimension: [% of GCT]), and $Limp$ (dimension: [% of GCT]), can be calculated following the approaches of classical gait analysis [13], [23]. Spatial parameters, such as Stride Length/Velocity (SL/SV, dimension: [% of patient's height]/[% of patient's height/s]) and Step Length/Velocity ($StepL/StepV$, dimension: [% of patient's height]/[% of patient's height/s]), are estimated modeling gait as an inverted pendulum and using the vertical displacement (h) of the trunk and the leg length (ℓ) to compute the forward distance (D) traveled at each step [20]. Important information about the mobility of the lower limbs is extracted using the gyroscopes of the Shimmer devices placed on the thighs. Integrating the angular rate measured by the gyroscopes, the instantaneous inclination of the thighs (θ) can be estimated. The angular amplitude of a thigh's flexion/extension movement, denoted as thigh range of rotation ($Thigh\ RoR$, dimension: [deg]), is then computed considering the maximum and the minimum inclination values in each gait cycle. Moreover, we calculate and analyze the unbiased autocorrelation (A_{unbiased}) associated with the acceleration signals recorded with the trunk-mounted IMU to obtain, in a simple way, additional information about the regularity and periodicity of patients' walking patterns, such as the Step Regularity (R_{step}, adimensional) and the step/stride Symmetry (S, adimensional) [24]. Finally, similarly to the LA case, in the frequency domain we compute the spectra of the three components of the trunk acceleration, denoted, respectively, as $X_{a_{\mathrm{vert}}}$, $X_{a_{\mathrm{z}}}$, and $X_{a_{\mathrm{y}}}$. The power associated with each spectrum is then calculated and their sum, denoted as P_{sum} (adimensional), is a relevant kinematic feature.

For ease of visualization and according to the results obtained in [10], in the following, only a reduced set of features is considered. In particular, the parameters with right and left

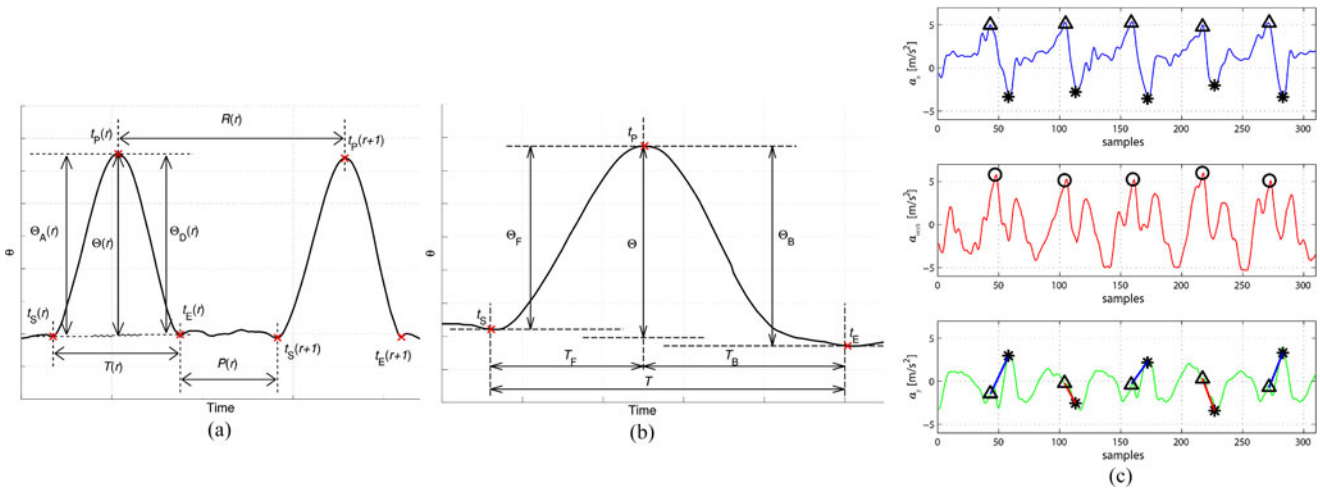


Fig. 3. Typical patterns in (a) the inclination signal θ of one thigh during the LA task (for two consecutive repetitions), (b) the torso inclination signal θ during the S2S task, and (c) the trunk acceleration signals during the G task. In (a) and (b), the fundamental time events are denoted with red crosses and the intuitive representations of the most relevant features are also shown. In (c), the circles in linear vertical acceleration a_{vert} identify the searching region in the frontal acceleration a_{z}, within which the HS and TO events, denoted, respectively, with triangles and asterisks, are identified. In the mediolateral acceleration a_{y}, HS points are connected with consecutive TO points by a line whose slope allows to discriminate left leg (blue line, positive slope) from right leg features (red line, negative slope).

components are replaced with the arithmetic average of the two values. The reduced set of features for the G task is the following: $\{GCT_{\mathrm{mean}}, ST_{\mathrm{mean}}, DS, Limp, SL, StepV_{\mathrm{mean}}, C, R_{\mathrm{step}}, S, Thigh\ RoR_{\mathrm{mean}}, P_{\mathrm{sum}}\}$.

B. Automatic Classification Approach

As mentioned in Section I, a key tool behind this study is the design and implementation of a unique system for the automatic UPDRS scoring of the LA, S2S, and G tasks, based on the assessment of the relevant kinematic features outlined in Section IV-A. In our previous works [7], [8], [10], although each task is analyzed singularly, the same approach for data processing, automatic classification, and performance analysis is used. In the following, a brief description of the used methods is presented.

1) Principal Component Analysis: In order to reduce the features' dimensionality and redundancy, while retaining most of the information content of the original data, principal component analysis (PCA) is applied on the collections of kinematic features defined for each task. Before applying PCA, the original data are first centered at their means (which are set equal to 0) and rescaled to have unit standard deviation. For the automatic classification procedure presented in the following, both original and "PCA-projected" data will be considered as input.

2) Classification Algorithms and Performance Analysis: The automatic UPDRS scoring system relies on the use of the consensus UPDRS score ($\mathrm{UPDRS}_{\mathrm{Mean}}$), i.e., on the arithmetic average of the UPDRS scores assigned by the three neurologists to each trial. Three well-known classifiers have been considered: nearest centroid classifiers (NCC), k nearest neighbors (kNN), and support vector machine (SVM) [25]. The chosen classification algorithms have different characteristics in terms of complexity and effectiveness and, thus, represent a good starting point to evaluate the feasibility of the proposed system. In order to avoid biasing the classification performance and considering the amount of data available, the leave-one-out cross-validation method is chosen. In particular, each point of both the original and "PCA-projected" dataset is used, in turn, as a new (unknown) point to be classified, while the remaining samples are used to train the classifiers. The result of the classification procedure is an estimated UPDRS value, generally denoted as $\widehat{u}_M$, for every trial of each task. The system performance is evaluated considering the absolute[3] classification error e_M, defined as follows:

$$e_M \triangleq |\widehat{u}_M - u_M|$$

where u_M is the value of $\mathrm{UPDRS}_{\mathrm{Mean}}$ for the considered trial ($u_M, \widehat{u}_M \in \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$).

In [7], [8], and [10], the automatic UPDRS scoring systems designed for the three tasks of interest have been exhaustively tested in order to determine the configurations, in terms of features combinations and system parameters, which allow us to achieve the best classification performance. For each task, the cumulative distribution function (CDF) of the error e_M is computed considering the results of the classification procedure obtained using, as inputs for the three classifiers (NCC, kNN, and SVM): 1) all the possible combinations of kinematic features with original data; and 2) increasing number of principal components (up to the number of features in the original dataset) when PCA-projected data are considered. Furthermore, when kNN is used as classification algorithm, values of k between 1 and 10 are used. The area under the curve (AuC) of the CDF of e_M is selected as a representative performance optimization metric, since maximizing this value corresponds to minimizing

[3]The absolute value of the classification error is considered because we are interested in quantifying the absolute deviation between automatically estimated and neurologist-assigned UPDRS scores.

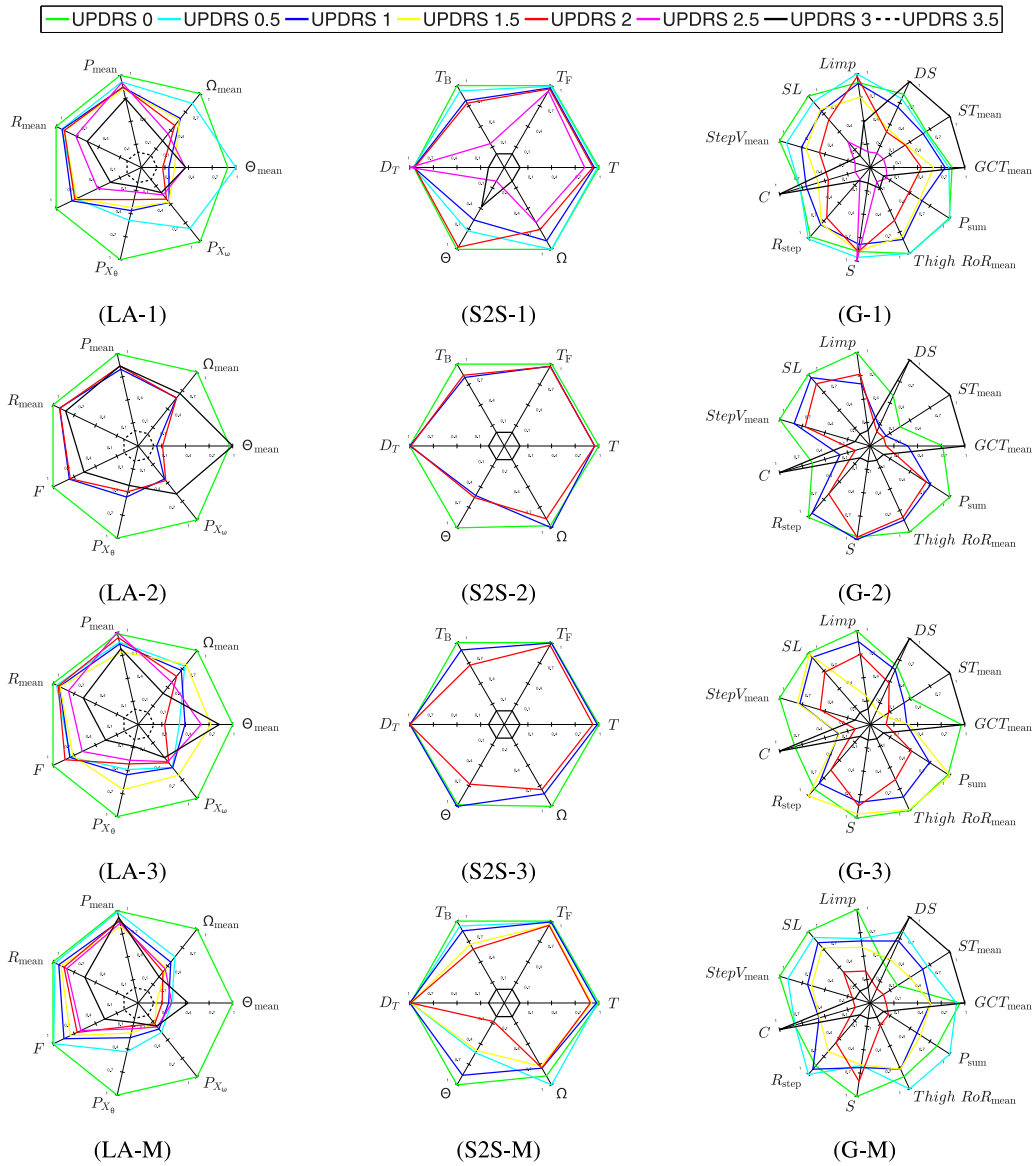


Fig. 4. Radar plots of the average normalized features grouped by UPDRS class in the LA, S2S, and G tasks considering the UPDRS scores by neurologist 1 [cases (LA-1), (S2S-1), (G-1)], neurologist 2 [cases (LA-2), (S2S-2), (G-2)], neurologist 3 [cases (LA-3), (S2S-3), (G-3)], and the $\mathrm{UPDRS}_{\mathrm{Mean}}$ [cases (LA-M), (S2S-M), (G-M)].

the overall absolute classification error. Among all the CDFs obtained for all considered parameters' combinations, those which maximize the AuC are selected to determine the system configuration which achieves the best classification performance.

V. Results

A. Kinematic Characterization

As already discussed in [7], [8], and [10], in each task, a strong relationship can be identified between some of the extracted kinematic features and the UPDRS score assigned by the neurologists. In Fig. 4, the average values, over all trials belonging to each UPDRS class (from 0 to 4), of the most relevant features for the LA, S2S, and G tasks are shown, through radar plots, considering the UPDRS scores given by the three neurologists and the $\mathrm{UPDRS}_{\mathrm{Mean}}$. Each parameter has been normalized and rescaled in order to assume a value between 0 and 1, where 0 represents the worst case and 1 the best case—best and worst are feature-specific. As expected, it can be observed that, beyond the interneurologist variability, in each task, the overall motor performance of the patients, in terms of its associated kinematic parameters, tend to worsen for increasing UPDRS scores. The values of the parameters corresponding to the UPDRS class 0, indeed, achieve the best performance in almost all the considered features and in every task, whereas the values belonging to higher UPDRS classes (from 2.5 to 3.5) tend to lie in the region of the plot near the origin, which is thus associated with the worst performance. In order to show how much every single feature is related to the UPDRS score, we compute the Pearson's correlation coefficient (denoted also as r-value)

between the kinematic parameters and the $\text{UPDRS}_{\text{Mean}}$ values of each task. In the last column of Table II, the r-values for all the most relevant features are shown, highlighting in bold those which have the highest correlation with the $\text{UPDRS}_{\text{Mean}}$ score of the considered task. We note that correlations with the corresponding p-value[4] higher than 0.05 are considered as nonsignificant (n.s.). Furthermore, the sign of the correlation coefficients makes it possible to discern if the value of the considered feature increases (positive sign) or decreases (negative sign) for increasing UPDRS score.

For the LA task, the average values per UPDRS class is obtained considering both RLA and LLA trials. Looking at Fig. 4 (LA-1), (LA-2), (LA-3), and (LA-M), it can be observed that the decreasing trend for increasing UPDRS score is evident in all the considered features. Values belonging to the extremes UPDRS classes (e.g., UPDRS scores equal to 0 or 3/3.5) are clearly separated from the others, while values associated with the intermediate UPDRS classes (e.g., UPDRS scores equal to 1, 1.5, or 2) tend to overlap, in some cases, in the same region of the plot. This behavior is consistent with clinical evaluations of PD patients, when distinguishing between intermediate levels of impairments may be difficult. Considering the correlations between LA features and UPDRS score, the parameter which has the highest (absolute) r-value is Ω_{mean}.

Similarly to the LA case, in the S2S task, the performance degradation for increasing UPDRS score is evident, as shown in Fig. 4 (S2S-1), (S2S-2), (S2S-3), and (S2S-M). The parameters for which this trend is clearer are those related to the duration of the S2S single rising, namely, T, T_F, and T_B. As expected, these parameters achieve the highest correlation values with respect to the UPDRS score and are those that better characterize the S2S task.

In the G task, the trends of the features are more difficult to interpret due to the higher complexity of the body movements involved in this exercise. Considering the temporal parameters, such as GCT, ST, DS, and $Limp$, the performance decreases for increasing UPDRS score. Patients with gait impairments, indeed, tend to walk more slowly than normally walking subjects, increasing the duration of GCT and the permanence in the DS phase. This normally corresponds to lower values of C, since the number of steps, which can be performed by a patient in a minute reduces. Nevertheless, in Fig. 4 (G-1), (G-2), (G-3), and (G-M), it can be observed that the "best" performance in some of the temporal features is achieved by the average values associated with UPDRS class 3. This is due to the fact that PD patients who present *festinating gait*, i.e., a gait pattern alteration typical of Parkinsonians, characterized by a quickening and shortening of normal strides, perform short steps with a very high cadence, thus leading to values in temporal parameters and cadence that may be interpreted as "good" even for high UPDRS scores. Spatial parameters, flexion/extension excursions of the thighs, and step regularity show similar behaviors, with a clear decreasing trend for increasing UPDRS values. This result is consistent with clinical observations of Parkinsonian walking, in which patients with increasing gait impairments perform shorter steps, with reduced velocity and regularity, revealing, in general, a more limited movement range in lower limbs [27]. This observation is also confirmed by the good correlation values (absolute values approximately equal to 0.6) between kinematic parameters, such as SL, $StepV_{\text{mean}}$, R_{step}, and $Thigh\ RoR_{\text{mean}}$, and the $\text{UPDRS}_{\text{Mean}}$ score of the G task. The S parameter maintains a low variability across all UPDRS values, except for the subjects with UPDRS score equal to 3: this seems to be more related to the single subject walking characteristics than to the entire scoring cluster. Finally, the feature P_{sum}, which is representative of the overall "power" associated with the patient's movements during gait, clearly decreases monotonically from UPDRS 0 to UPDRS 3, showing also a moderate correlation (r-value equal to -0.43) with the UPDRS score.

[4]We recall that the p-value represents the probability that the observed differences, in the sample data which are being tested, are due to random sampling errors and not to true differences between populations [26].

TABLE III
COMBINATION OF PARAMETERS CORRESPONDING TO THE BEST PERFORMANCE OF THE AUTOMATIC UPDRS SCORING SYSTEM FOR EACH TASK

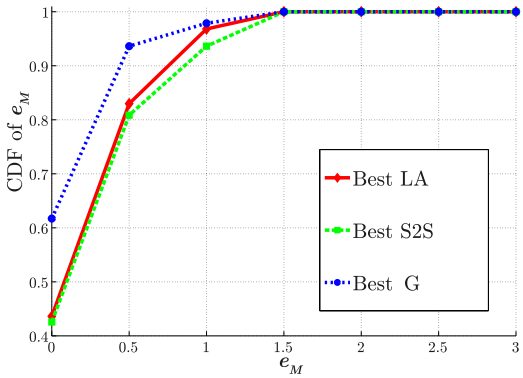
Task	Classifier	Set of features
LA	kNN ($k = 3$)	$\{\Omega_{\text{mean}}, F\}$
S2S	kNN ($k = 3$)	T
G	kNN ($k = 6$)	$\{DS, R_{\text{step}}, Thigh\ RoR_{\text{mean}}\}$

B. Automatic Detection

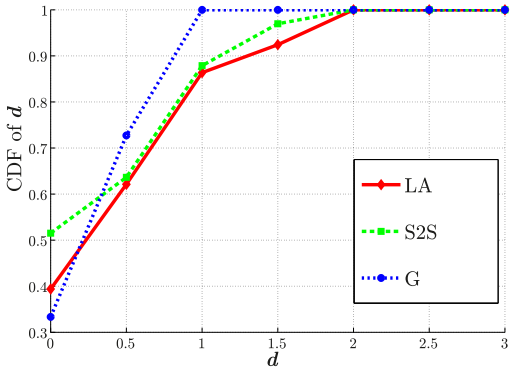
As anticipated in Section IV-B2, an exhaustive testing of the automatic UPDRS scoring system for each task has been carried out in previous papers [7], [8], [10] in order to find the parametric configuration which allows to achieve, for each task, the best classification performance, i.e., which maximizes the AuC of the CDF of the error e_M. In Table III, the optimal system configuration, including the best classification algorithm and the best combination of features, is shown for each task. The observed results show that in all tasks the best classifier is kNN, with k set to 3 (LA and S2S) or 6 (G). Furthermore, it can be observed that the number of features associated with the best performance increases for increasing complexity of the task movement patterns.

In Fig. 5(a), the CDFs corresponding to the optimized parametric configurations of the automatic UPDRS scoring system in each task are shown. The *accuracy* of the automatic system (corresponding to CDFs' values at $e = 0$) ranges from approximately 43% in the LA and S2S tasks to 62% in the G task. The percentage of trials classified with $e \leq 0.5$ is over 81% in all the considered tasks (LA: 83%; S2S: 81%; and G: 94%) while more than 94% of the samples are classified with $e \leq 1$ (LA: 97%; S2S: 94%; and G: 98%). Moreover, it can be observed that the classification error is never greater than 1.5.

To better characterize the classification performance with more details, in Tables IV–VI, the confusion matrices associated with the automatic scoring procedures of the LA, S2S, and G tasks are shown, respectively. It is easy to observe that the automatic system tends, in all cases, to bias the estimated



(a)



(b)

Fig. 5. CDFs of (a) the automatic classification error e_M for the best performance achieved in each task and (b) the (absolute) difference d in the UPDRS scoring between neurologist 1 and 2.

TABLE IV
CONFUSION MATRIX FOR THE LA TASK AUTOMATIC UPDRS SCORING PROCEDURE

[%]	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
0	**0**	50	50	0	0	0	0	0
0.5	8	**54**	23	15	0	0	0	0
1	0	23	**18**	53	6	0	0	0
1.5	0	5	16	**65**	11	3	0	0
2	0	17	0	58	**25**	0	0	0
2.5	0	0	0	50	10	**40**	0	0
3	0	0	0	50	0	50	**0**	0
3.5	0	0	0	0	0	100	0	**0**

TABLE V
CONFUSION MATRIX FOR THE S2S TASK AUTOMATIC UPDRS SCORING PROCEDURE

[%]	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
0	**100**	0	0	0	0	0	0	0
0.5	47	**31**	1	1	0	0	0	0
1	40	40	**20**	0	0	0	0	0
1.5	34	0	22	**22**	22	0	0	0
2	34	0	0	66	**0**	0	0	0
2.5	-	-	-	-	-	**-**	-	-
3	-	-	-	-	100	0	**0**	0
3.5	-	-	-	-	-	-	-	**-**

TABLE VI
CONFUSION MATRIX FOR THE G TASK AUTOMATIC UPDRS SCORING PROCEDURE

[%]	**0**	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**
0	**0**	0	100	0	0	0	0	0
0.5	0	**45**	55	0	0	0	0	0
1	0	0	**78**	22	0	0	0	0
1.5	0	6	25	**69**	0	0	0	0
2	0	0	0	100	**0**	0	0	0
2.5	-	-	-	-	-	**-**	-	-
3	0	0	0	100	0	0	**0**	0
3.5	-	-	-	-	-	-	-	**-**

UPDRS values around a dominant class, in accordance to the Gaussian-like distribution of the $\text{UPDRS}_{\text{Mean}}$ score observed in Section III-B. The UPDRS classes with a small number of samples are almost "ignored" by the classifier and the samples associated with them are labeled with a UPDRS value nearer to the dominant class, thus determining a general underestimation of actual UPDRS score. This behavior is a clear consequence of the non-homogeneity of the UPDRS evaluations of the various neurologists. From the presented confusion matrices, other per-class performance indexes, such as the *precision*, the *sensitivity*, and the *specificity*, can be calculated—they are not shown here due to lack of space. We only remark that the average values (across all the UPDRS classes) of these indexes are equal to (*precision*, *sensitivity*, *specificity*) 34.55%, 25.17%, and 84.52% in the LA task, 28.00%, 25.63%, and 77.51% in the S2S task, and 66.48%, 31.83%, and 88.03% in the G task.

Finally, in order to evaluate the operational correctness of our automatic scoring system, we compare the achieved performance with the inter-rater variability of the neurologists in the UPDRS task assessment. In Fig. 5(b), the CDFs of the (absolute) difference d between the UPDRS scores assigned by neurologists 1 and 2 in the three tasks are shown. The agreement in the evaluations ranges from 33% in the G task to 52% in the S2S task, whereas the difference in the UPDRS scores between the two neurologists is lower than or equal to 1 in approximately 90% of the trials (100% for the G task). Similar results have been obtained from the comparison of the evaluations by neurologists 1 and 3 and by neurologists 2 and 3—the corresponding CDFs are not shown here for lack of space. Comparing Fig. 5(a) and (b), it is possible to observe very similar trends in the CDFs of the estimation error e_M and the difference of neurologists' evaluations d. It can thus be concluded that the variability in the UPDRS scoring between the automatic system and the neurologists seems to be comparable with the inter-rater variability between clinicians. In other words, the proposed automatic classification system is "accurate" enough to mimic the evaluation performance of medical personnel.

Obviously, a larger set of patients, a more uniform distribution of the patients in all the UPDRS classes, and additional evaluations by more neurologists would make the proposed system performance analysis more meaningful from a statistical perspective. However, this goes beyond the scope of the paper,

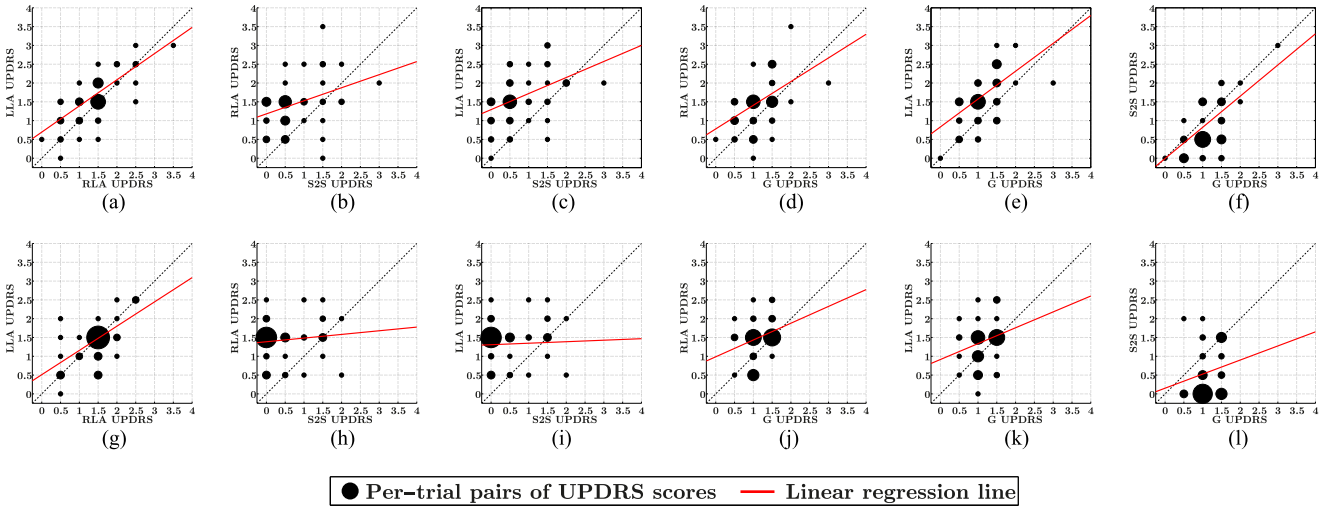


Fig. 6. Comparison between UPDRS values assigned to each trial for different pairs of UPDRS tasks considering: in the upper row [(a)–(f)], the $\text{UPDRS}_{\text{Mean}}$ scores; in the lower row [(g)–(l)], the UPDRS values estimated through the automatic scoring system. Each point corresponds to a "cluster" of trials labeled with the same pair of UPDRS values and its size is proportional to the number of trials belonging to the same cluster. (a) RLA versus LLA. (b) S2S versus RLA. (c) S2S versus LLA. (d) G versus RLA. (e) G versus LLA. (f) G versus S2S. (g) RLA versus LLA auto. (h) S2S versus RLA auto. (i) S2S versus LLA auto. (j) G versus RLA auto. (k) G versus LLA auto. (l) G versus S2S auto.

which focuses on proposing a novel approach, rather than an exhaustive medical investigation.

C. Correlations

In Section V-A, we have investigated the correlations between the most relevant kinematic features in each task and the UPDRS scores assigned by the neurologists. Since we are investigating the feasibility of a single system able to automatically assess different motor tasks of Parkinsonians and to automatically assign UPDRS scores to patients' trials, we now focus on the analysis of the correlations between the UPDRS evaluation carried out by doctors and the one performed by our automatic UPDRS scoring system. Looking at Fig. 2, the differences in the distribution of the UPDRS values in the three tasks, performed by the same group of patients, can be easily distinguished in the (single) UPDRS scores assigned by the neurologists and the $\text{UPDRS}_{\text{Mean}}$ score. As expected, this means that the performance achieved by patients in the three tasks and, consequently, the UPDRS evaluations by doctors, may vary significantly from one task to another. In Fig. 6, a comparison between the $\text{UPDRS}_{\text{Mean}}$ scores [cases from (a) to (f)] and the UPDRS values assigned by the automatic system [cases from (g) to (l)] is shown considering pairs of tasks.[5] Each trial is labeled with a pair of UPDRS values, corresponding to the UPDRS scores assigned to it in the two considered tasks. In each subplot, the "clusters" of trials labeled with the same pair of UPDRS scores and, thus, overlapping on the same portion of the plane, are shown as black points, whose size is proportional to the number of trials belonging to the cluster. Moreover, the linear regression line, i.e., the best-fitting straight line obtained with the least squares method, is also shown. For all the considered pairs of tasks, an increasing trend in UPDRS scores for both tasks can be observed, although for each UPDRS value assigned to one task, several UPDRS values can be assigned to the other task.

For some pairs of tasks, such as those shown in Fig. 6(a) and (d)–(f), the linear relationship between the UPDRS scores is evident, and consequently, the regression line lies near the diagonal. In these cases, the values assigned in both tasks, in fact, are likely to be similar, indicating both comparable difficulties experimented by patients in the two tasks and a uniform metric used by the neurologists to assess them. In other cases, especially those including the S2S task, it can be observed that, while in one task (either LA or G), the trials are labeled with increasing UPDRS scores, in the S2S task, the patient is still able to achieve a good performance, and consequently, the trials are labeled with a UPDRS equal to 0 or 0.5. This leads to an accumulation of points near the 0 and 0.5 UPDRS classes for the S2S task, which also influences the slope of the linear regression line, making it "flatter." This behavior is even more visible considering the automatically estimated UPDRS scores [cases from (g) to (l)]. It can been observed that the linear relationship between tasks is always weaker than in the neurologist-assessed cases. This is due to the fact that the automatic scoring system, as anticipated in Section V-B, tends to slightly underestimate the values of the UPDRS scores with respect to those assigned by doctors, increasing the concentration of the scores around the dominant UPDRS class of each task.

To reinforce the considerations made observing the pairs of tasks in Fig. 6, we now analyze numerically the correlation between the considered tasks. We also introduce an aggregate UPDRS score, denoted as $\text{UPDRS}_{\text{Total}}$, given by the sum of the $\text{UPDRS}_{\text{Mean}}$ scores assigned to the RLA, LLA, S2S, and G tasks. The summation of the UPDRS scores assigned to different combinations of tasks [13], [17], [28], is a common practice used in the literature to predict the overall functional capabilities

[5]For the LA task, comparisons with both RLA and LLA UPDRS scores are shown separately.

TABLE VII
CORRELATIONS BETWEEN $\text{UPDRS}_{\text{Mean}}$ SCORES

	Total	RLA	LLA	S2S	G
Total	1.00				
RLA	0.82	1.00			
LLA	0.87	0.75	1.00		
S2S	0.76	0.37	0.46	1.00	
G	0.82	0.51	0.60	0.66	1.00

TABLE VIII
CORRELATIONS BETWEEN UPDRS SCORES ASSIGNED BY THE AUTOMATIC UPDRS SCORING SYSTEM

	Total	RLA	LLA	S2S	G
Total	1.00				
RLA	0.77	1.00			
LLA	0.74	0.63	1.00		
S2S	0.58	n.s.	n.s.	1.00	
G	0.54	0.29	n.s.	n.s.	1.00

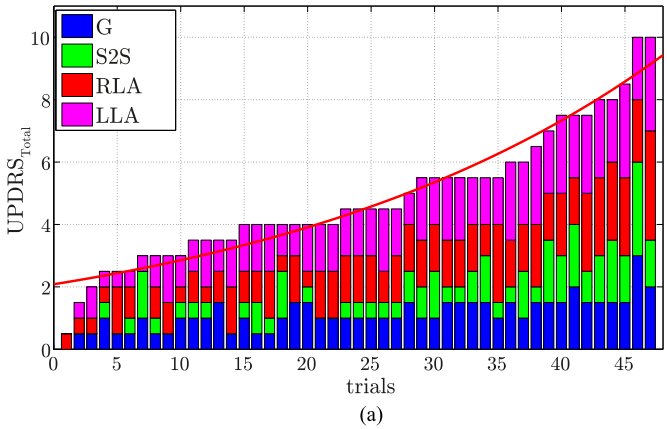
of PD patients. From our viewpoint, $\text{UPDRS}_{\text{Total}}$ is a concise parameter representing the overall level of motor impairments of a patient in the considered tasks.

In Table VII, the correlations between the $\text{UPDRS}_{\text{Mean}}$ scores in the various tasks and between them and $\text{UPDRS}_{\text{Total}}$ are shown. As in Section V-A, the Pearson's correlation coefficient (i.e., the r-value) is used, considering as significant only the parameters with associated p-value ≤ 0.05. It can be observed that the correlation between the UPDRS scores of each task and the aggregate score is high (from 0.76 to 0.87), indicating that $\text{UPDRS}_{\text{Total}}$ significantly represents, in a concise way, the motor performance level measured by each task. The correlations between pairs of tasks, instead, range from 0.75 (between RLA and LLA, which are likely to be strongly correlated since they refer to the same exercise) to 0.37 between RLA and S2S—this is representative of the poor correlation between those tasks. The latter result is expected because, from a medical viewpoint, each UPDRS task aims to assess a specific PD symptom and the "performance" achieved by patients may vary consistently from one exercise to another [17], [29]. However, in some cases, such as in the comparison between S2S and G UPDRS scores, the r-value is still relevant because both tasks refer to the same "macrogroup" of PD symptoms (in this case to the "Gait/Posture" group) and may share some characteristics [17].

In the same way, in Table VIII, the correlation values (obtained by pairwise comparisons of the UPDRS scores estimated with the automatic UPDRS scoring system in each task, and the associated $\text{UPDRS}_{\text{Total}}$) are presented. The obtained results still show relatively high r-values (although slightly lower the those of the neurologist-assessed cases) between tasks and the $\text{UPDRS}_{\text{Total}}$ score. Also, the correlation between RLA and LLA remains high. For all the other pairs of tasks, instead, the correlations become nonsignificant. This is due to the fact that, as observed in Fig. 6(h)–(l), the automatically estimated UPDRS scores are often underestimated and tend to overlap in correspondence to the dominant UPDRS class of each task, making

TABLE IX
CORRELATIONS BETWEEN $\text{UPDRS}_{\text{Mean}}$ SCORES AND AUTOMATICALLY ESTIMATED UPDRS SCORES

Total	RLA	LLA	S2S	G
0.79	0.74	0.54	0.55	0.60



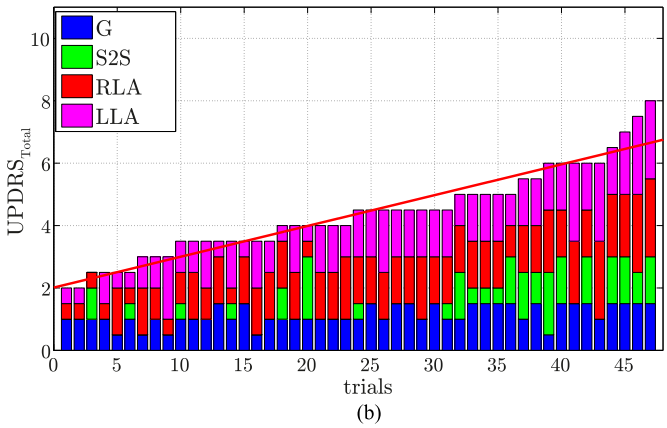


Fig. 7. Bar plots representing the values of $\text{UPDRS}_{\text{Total}}$ for all the trials considering (a) the associated $\text{UPDRS}_{\text{Mean}}$ scores and (b) the UPDRS values assigned by the automatic scoring system. Contributes by each task are shown with different colors.

the linear dependence between the UPDRS values in the pairs of tasks weaker.

For completeness, Table IX shows the correlations between the neurologist-assigned and the automatically assigned UPDRS scores. In $\text{UPDRS}_{\text{Total}}$ and RLA cases, the correlation is high (0.79 and 0.74, respectively), while in the LLA, S2S, and G tasks, lower values are observed (0.54, 0.55, and 0.60, respectively), indicating a higher sensitivity to classification errors.

Finally, in Fig. 7, the $\text{UPDRS}_{\text{Total}}$ values, calculated considering both (a) the $\text{UPDRS}_{\text{Mean}}$ and (b) the trials' scoring by the automatic system, are shown. The trials are ordered in ascending order to highlight the data trend for increasing values of $\text{UPDRS}_{\text{Total}}$. The contributions of the UPDRS scores assigned to the single tasks, for each trial, are shown with different colors. The (red) exponential curves, obtained by minimum mean square error fitting, represent smoothed versions of the aggregate scores' trends. First of all, it can be noticed that, as mentioned

TABLE X
REDUCTIONS IN THE AUC OF THE CDF OF e_M COMPARING THE VARIOUS PARAMETRIC CONFIGURATIONS WITH THE OPTIMAL ONE

	Method			
Number of features or principal comp.	**kNN**	**kNN-PCA**	**SVM**	**SVM-PCA**
		LA		
1	−2.79%	−5.20%	−3.76%	−10.99%
2	**Best LA**	−5.01%	−5.31%	−11.57%
3	−0.29%	−7.23%	−4.82%	−12.72%
4	−1.25%	−5.49%	−5.59%	−14.27%
5	−1.15%	−4.53%	−6.07%	−14.94%
		S2S		
1	**Best S2S**	−0.19%	−2.92%	−3.50%
2	−0.77%	−2.92%	−4.28%	−4.48%
3	−0.97%	−1.94%	−2.92%	−3.31%
4	−1.55%	−1.55%	−3.11%	−5.84%
5	−1.36%	−1.55%	−4.67%	−3.70%
		G		
1	−1.11%	−4.46%	−6.69%	−8.36%
2	−0.37%	−4.08%	−5.20%	−7.06%
3	**Best G**	−4.83%	−4.27%	−9.47%
4	−0.19%	−3.71%	−3.71%	−7.80%
5	−0.37%	−4.27%	−4.08%	−8.00%

above, the automatic classification system tends to assign UPDRS scores in accordance to the dominant UPDRS class of each task: for the G task, indeed, the majority of the scores is equal to 1; for the S2S case, it is equal to 0; and for the LA task, most of the trials are labeled with UPDRS scores equal to 1.5. This behavior leads to generally lower values of $\mathrm{UPDRS}_{\mathrm{Total}}$ and to a more reduced variability in the aggregate score for the automatic system, which is also emphasized by a lower slope of exponential fitting curve in Fig. 7(b). Another important observation regards the contribution given by the S2S scores to $\mathrm{UPDRS}_{\mathrm{Total}}$: it can be observed that it is almost negligible (between 0 and 0.5 in 90% of the trials) for aggregate scores lower than 5, especially when the automatic system is used. These experimental observations confirm the effectiveness of the LA and G tasks in representing the progression of motor impairments in PD patients and also highlight the "nonchallenging" nature of the S2S task for patients with mild symptoms. Nevertheless, the UPDRS score assigned to the S2S task becomes very important to distinguish between Parkinsonians with moderate and severe motor complications, when the $\mathrm{UPDRS}_{\mathrm{Total}}$ is used as evaluation metric.

VI. DISCUSSION

A. On the Performance of the Proposed Automatic Classification System

The results obtained in Section V-A indicate clearly that, in each task, some of the extracted kinematic features are strongly related to the UPDRS scores. The parameters that have turned out to be the most significant are the angular speed of execution (Ω) for the LA task, the total duration (T) for the S2S task, and the stride length (SL) for the G task.

Regarding the automatic UPDRS scoring system, the results presented in Section V-B have highlighted that the best classification performance is achieved using kNN as the classifier on the selected kinematic features, using an increasing number of features and increasing values of k for increasing complexity of the tasks. The accuracy of the system ranges from 43% (in the LA and S2S tasks) to 62% (in the G task), but the classification error is lower than or equal to 1 in more than 94% of the cases. The comparison between the evaluation error of the automatic system and the interrater variability of the neurologists has shown similar performance trends, allowing us to consider the accuracy of the proposed UPDRS scoring system acceptable. Nevertheless, the automatic system tends to underestimate the actual UPDRS scores and to concentrate the predicted UPDRS values in correspondence to dominant UPDRS classes. This is not an intrinsic limitation of the proposed approach. In fact, it could be overcome by considering a statistically more significant dataset (i.e., a larger set of patients, with a more homogeneous distribution across all UPDRS classes). In addition, increasing the number of neurologists involved in the evaluation could reduce the bias in the assessment of UPDRS motor tasks.

In Section V-C, the comparative analysis of the correlations suggests that the motor performance of PD patients may vary consistently between different tasks, and thus, the associated correlations may range from low (0.37, poor correlation) to high (0.75, good correlation) values. The correlation between UPDRS scores in distinct tasks becomes weaker (almost always negligible) when the automatic system is considered. This is another consequence of the fact that the automatic system tends to bias the evaluation around the dominant UPDRS classes. However, this result complies with findings in the medical literature [17], according to which the correlation between different tasks is likely to be poor or moderate because each task has been defined with the aim to evaluate a specific aspect or symptom of the PD.

Finally, our results have shown a good correlation between $\mathrm{UPDRS}_{\mathrm{Total}}$ and all the UPDRS scores of all the tasks (slightly lower correlations for the automatically assessed tasks). This concise index can provide neurologists useful information about the overall condition and the functional capabilities of patients, integrating the evaluations made considering the single UPDRS scores in each task. The contribution of the S2S task in the aggregate score, for example, seems to be significant to distinguish patients with slight and mild symptoms from those who manifest moderate or severe impairments. This observation can be related to the characteristics of the movement typical of the S2S task, which involves the simultaneous activation of several control mechanisms (visual, posture, and balance) and worsens with the progress of the disease. A further investigation on the role of the S2S task goes beyond the scope of this work.

B. Application to Telerehabilitation

The characteristics of the designed system make it suitable for real applications in the e-health scenario, such as telemedicine systems for remote monitoring of PD. The objective recording of motor fluctuations in a home environment throughout the day, unlike a time-limited and "randomly timed" clinical evaluation in an out-patient environment, could provide more reliable information to neurologists and allow a more accurate assessment and management of the symptoms. These goals could be

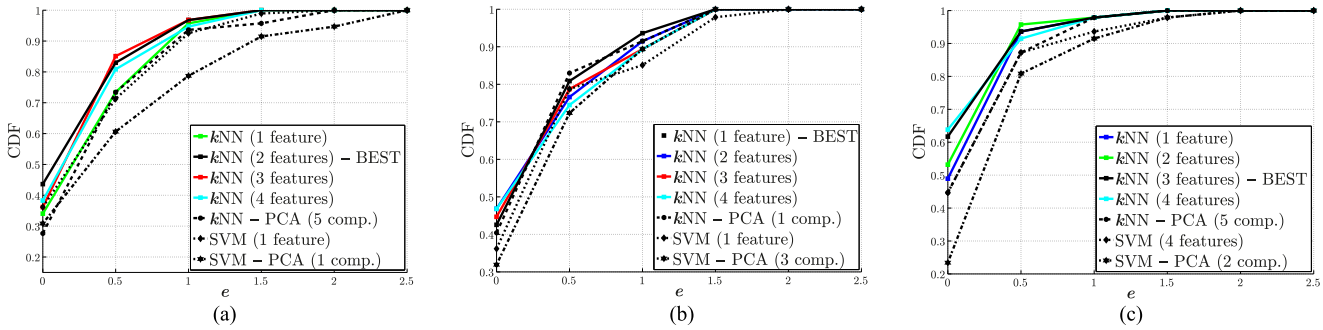


Fig. 8. CDFs associated with different system configurations in the (a) LA, (b) S2S, and (c) G tasks. The best CDF is represented with a solid black line.

achieved, for example, implementing a software which guides patients in performing different activities, such as the UPDRS tasks for the evaluation of the PD symptoms or rehabilitation exercises in the comfort of their own home and perhaps several times per day. The information about the patient's performance, acquired by the BSN, could be evaluated automatically or sent to the neurologist for further analysis and should be supported by a concurrent video acquisitions which clinicians can check to ensure that the exercises are performed correctly. With this expedient, the supervision of a movement disorder specialist would not be required, although the assistance of a relative or a nurse could be useful to check that each task is executed safely. A tool with these functionalities may bring several advantages 1) to patients, who would feel more comfortable and motivated to do their exercises in a familiar environment and would save time and money by avoiding to go to an ambulatory for each visit, and 2) to clinicians and doctors, who could assist a larger number of subjects and rely on more accurate and up-to-date clinical pictures.

Since the e-health systems market is quite fragmented and poorly standardized, the best strategy for a practical implementation of the proposed PD monitoring application could be to integrate it into an existing e-health platform [30], [31], exploiting the APIs provided by the platform. The integration could be done, for example, at the *gateway* level (i.e., the level at which the data recorded by the sensing devices are collected, analyzed, and then forwarded through the web to the cloud-based core of the e-health platform), by adding a proper software component for the analysis of the data recorded by the IMU-based BSN. The processed data could then be managed according to the platform requirements, in the same way as the other data collected by the platform.

C. Efficient Implementation

In the previous sections, we have designed and tested our system in order to investigate its feasibility and performance, without considering constraints in terms of computational power or time consumption. In a real application scenario, as the one discussed in Section VI-B, the system would be subject to several limitations. An efficient design and implementation would be thus required in order to reduce the complexity and, consequently, the required computational resources, while still retaining the ability to achieve good (although not optimal) performance. To this end, we now evaluate and discuss the differences in performance between the optimal system configuration and other "suboptimal," but simpler, alternatives.

In Table X, the performance reduction (in percentage), obtained by comparing the AuC of the CDF of the error e_M associated with the best parametric configuration with the AuCs of "suboptimal" alternatives, are shown. The best configurations obtained for the three tasks, previously shown in Table III, are denoted in Table X as Best LA (AuC equal to 92.29%), Best S2S (AuC equal to 90.96%), and Best G (AuC equal to 95.39%) and highlighted in bold. Moreover, in Fig. 8, the CDFs associated with the kNN algorithm applied to an increasing number of features (from 1 to 4) of the original data and the best configurations obtained using 1) kNN and PCA, 2) SVM on the original dataset, and 3) SVM on the "PCA-projected" data, are shown to provide a visual counterpart of the data presented in Table X. As shown in Section V-B, it can be observed that the best classification algorithm is the kNN in all the considered tasks. This kind of classifier does not require an explicit training phase, as it keeps all the dataset points (true neurologist-based scores) to take decisions in the "online" phase. This so-called lazy learning approach, however, implies high memory consumption and computational power to check all the training set elements for each classification round. On the other hand, the SVM algorithm achieves a performance similar to (actually, slightly worse than) the one of the kNN method but with a more efficient learning procedure. In fact, SVM, after a more complex training phase, builds a compact classification model which can be used to simply identify the class of the test data. The relative reduction (in percentage), considering the best configuration in the SVM case for all the three tasks, is very limited, ranging from 2.92% (S2S task) to 3.76% (LA task).

Another possible strategy to lower the complexity of the automatic classification procedure is to reduce the dimensionality of the features dataset used as input for the classifiers. Looking at Table X, it can be observed that the performance reduction, considering only one or two features (original data) for the kNN algorithm, is almost always below 2.80%. In the SVM case as well, the performance degradation, with respect to the best SVM case, using only one feature is minimal. Dimensionality reduction in the presence of PCA tends to worsen the overall performance with both kNN and SVM methods. From these

considerations, it can be concluded that, although the best classification performance is achieved by the kNN classifier, the SVM classifier, applied to features of the original dataset, could be a more attractive choice for real-world applications, because of 1) higher efficiency (in terms of computational resources consumption) and 2) minimal classification accuracy reduction with respect to kNN.

VII. Conclusion

We have proposed an innovative approach for kinematic characterization of the LA, S2S, and G tasks through the same BSN with three nodes, together with automatic UPDRS assessment of the trials carried out by PD patients. Building on this unified approach, we investigated the intertask correlations and the correlation, per task, between the proposed automatic scoring system and the neurologists' scoring. The main findings of our analysis can be summarized as follows.

1) Different UPDRS tasks aims to assess distinct aspects of the disease and show poor or moderate correlations between each other, especially with the automatic classification system.
2) On the basis of the scores given by three expert neurologists, our results show that the performance of the proposed automatic classification system compares favorably with typical interrater variability.
3) The aggregate UPDRS score ($\mathrm{UPDRS}_{\mathrm{Total}}$) represents a good concise indicator of overall level of patient's motor impairments.
4) The UPDRS scores associated with the LA and G tasks are effective in representing the progression of motor impairments in PD patients and contribute to the $\mathrm{UPDRS}_{\mathrm{Total}}$ score proportionally to the level of difficulty experienced by the patient. The S2S task, instead, appears to be "non-challenging" for patients with light or mild symptoms, but its contribution in the aggregate UPDRS score becomes important in identification of the subjects with most severe impairments.

The integration of the proposed system in a real cloud-based e-health platform for the development of a telemedicine application for continuous monitoring of PD patients has then been discussed, with focus on a possible system architecture and possible strategies for an efficient implementation of the proposed functionalities.

The analyses and findings discussed in this paper can be further investigated to overcome their current limitations. In particular, the proposed automatic UPDRS scoring system can be improved by considering a larger and, thus, statistically more relevant dataset, together with clinical evaluations of additional neurologists. Moreover, a motor performance evaluation of PD patients in UPDRS tasks considering separately ON and OFF states represents an interesting research direction.

Acknowledgment

The anonymous Reviewers provided relevant feedback, which led to a significant improvement of this work. The authors would like to thank Dr. L. G. Pradotto (Istituto Auxologico Italiano IRCCS, Piancavallo (VB), Italy) for his contribution in the clinical evaluation of the subjects considered in this research. They are also grateful to Prof. K. Friedl (University of California, San Francisco, CA, USA) for carefully revising the final version of this manuscript and for his valuable suggestions.

References

[1] C. G. Goetz, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, G. T. Stebbins, M. B. Stern, B. C. Tilley, R. Dodel, B. Dubois, R. Holloway, J. Jankovic, J. Kulisevsky, A. E. Lang, A. Lees, S. Leurgans, P. A. LeWitt, D. Nyenhuis, C. W. Olanow, O. Rascol, A. Schrag, J. A. Teresi, J. J. V. Hilten, and N. LaPelle, "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, Jan. 2007.

[2] B. Post, M. P. Merkus, R. M. de Bie, R. J. Haan, and J. D. Speelman, "Unified Parkinson's disease rating scale motor examination: Are ratings of nurses, residents in neurology, and movement disorders specialists interchangeable?" *Movement Disorders*, vol. 20, pp. 1577–1584, Dec. 2005.

[3] J. E. Ahlskog and R. J. Uitti, "Rasagiline, parkinson neuroprotection, and delayed-start trials," *Neurology*, vol. 74, pp. 197–203, Jun. 2010.

[4] K. Kieburtz, "Issues in neuroprotection clinical trials in Parkinsons disease," *Neurology*, vol. 66, pp. 550–557, May 2006.

[5] R. A. Hauser, F. Deckers, and P. Lehert, "Parkinson's disease home diary: Further validation and implications for clinical trials," *Movement Disorders*, vol. 19, pp. 1409–1413, Dec. 2004.

[6] W. Maetzler, J. Domingos, K. Srulijes, J. J. Ferreira, and B. R. Bloem, "Quantitative wearable sensors for objective assessment of Parkinson's disease," *Movement Disorders*, vol. 28, pp. 1628–1637, Oct. 2013.

[7] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro, "Assigning UPDRS scores in the leg agility task of Parkinsonians: Can it be done through BSN-based kinematic variables?" *IEEE Internet Things J.*, vol. 2, no. 1, pp. 41–51, Feb. 2015.

[8] M. Giuberti, G. Ferrari, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro, "Automatic UPDRS evaluation in the sit-to-stand task of Parkinsonians: Kinematic analysis and comparative outlook on the leg agility task," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 3, pp. 803–814, May 2015.

[9] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro, "Low-complexity inertial sensor-based characterization of th UPDRS score in the gait task of Parkinsonians," presented at the 9th Int. Conf. Body Area Netw., London, U.K., Sep./Oct. 2014.

[10] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro. (2015, Aug.). Inertial BSN-based characterization and automatic UPDRS evaluation of the gait task of Parkinsonians," *IEEE Trans. Affective Comput.*. [Online]. Available: http://www.tlc.unipr.it/ferrari/TAFFC_Parisi_et_al_Feb2015.pdf

[11] D. A. Heldman, D. E. Filipkowski, D. E. Riley, C. M. Whitney, B. L. Walter, S. A. Gunzler, J. P.Giuffrida, and T. O. Mera, "Automated motion sensor quantification of gait and lower extremity bradykinesia," in *Proc. 34th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, San Diego, CA, USA, Aug. 2012 pp. 1956–1959.

[12] D. Giansanti, G. Maccioni, F. Benvenuti, and V. Macellari, "Inertial measurement units furnish accurate trunk trajectory reconstruction of the sit-to-stand manoeuvre in healthy subjects," *Med. Biol. Eng. Comput.*, vol. 45, no. 10, pp. 969–976, Oct. 2007.

[13] A. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, and K. Aminian, "Gait assessment in Parkinson's Disease: Toward an ambulatory system for long-term monitoring," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 8, pp. 1434–1443, Aug. 2004.

[14] S. T. Moore, H. G. MacDougall, J. Gracies, H. S. Cohen, and W. G. Ondo, "Long-term monitoring of gait in Parkinson's disease," *Gait Posture*, vol. 26, no. 2, pp. 200–207, Jul. 2007.

[15] A. Salarian, H. Russmann, C. Wider, P. R. Burkhard, F. J. G. Vingerhoets, and K. Aminian, "Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 2, pp. 313–322, Feb. 2007.

[16] J. Stochl, A. Boomsma, E. Ruzicka, H. Brozova, and P. Blahus, "On the structure of motor symptoms of Parkinson's disease," *Movement Disorders*, vol. 23, no. 9, pp. 1307–1312, 2008.

[17] S. D. Vassar, Y. M. Bordelon, R. D. Hays, N. Diaz, R. Rausch, C. Mao, and B. G. Vickrey, "Confirmatory factor analysis of the motor unified Parkinson's disease rating scale," *Parkinson's Disease*, vol. 2012, p. 719167, 2012.

[18] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, "SHIMMER—A wireless sensor platform for noninvasive biomedical research," *IEEE Sens. J.*, vol. 10, no. 9, pp. 1527–1534, Sep. 2010.

[19] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Zurich, Switzerland, Jun. 2011, pp. 1–7.

[20] W. Zijlstra and A. L. Hof, "Assessment of spatio-temporal gait parameters from trunk accelerations during human walking," *Gait Posture*, vol. 18, no. 2, pp. 1–10, Oct. 2003.

[21] A. Köse, A. Cereatti, and U. Della Croce, "Bilateral step length estimation using a single inertial measurement unit attached to the pelvis," *J. Neuroeng. Rehabil.*, vol. 9, pp. 1–9, Jan. 2012.

[22] J. Stamatakis, J. Ambroise, J. Crémers, H. Sharei, V. Delvaux, B. Macq, and G. Garraux, "Finger tapping clinimetric score prediction in Parkinson's disease using low-cost accelerometers," *Comput. Intell. Neurosci.*, vol. 2013, p. 717853, Jan. 2013.

[23] K. Aminian, B. Najafi, C. Büla, P. F. Leyvraz, and P. Robert, "Spatio-temporal parameters of gait measured by an ambulatory system using miniature gyroscopes," *J. Biomech.*, vol. 35, no. 5, pp. 689–699, May 2002.

[24] R. Moe-Nilssen and J. L. Helbostad, "Estimation of gait cycle characteristics by trunk accelerometry," *J. Biomech.*, vol. 37, pp. 121–126, Jan. 2004.

[25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis,*2nd ed. New York, NY, USA: Wiley-Interscience, 2000.

[26] R. A. Fisher, *Statistical Methods for Research Workers*. Edinburgh, U.K.: Oliver and Boyd, 1925.

[27] M. Svehlk, E. B. Zwick, G. Steinwender, W. E. Linhart, P. Schwingenschuh, P. Katschnig, E. Ott, and C. Enzinger, "Gait analysis in patients with Parkinson's disease off dopaminergic therapy," *Arch. Phys. Med. Rehabil.*, vol. 90, no. 11, pp. 1880–1886, 2009.

[28] J. Song, B. Fisher, G. Petzinger, A. Wu, J. Gordon, and G. J. Salem, "The relationships between the unified Parkinson's disease rating scale and lower extremity functional performance in persons with early-stage Parkinson's disease," *Neurorehabilitation Neural Repair*, vol. 23, no. 7, pp. 657–661, 2009.

[29] K. J. Brusse, S. Zimdars, K. R. Zalewski, and T. M. Steffen, "Testing functional performance in people with Parkinson disease," *Phys. Therapy*, vol. 85, pp. 134–141, Feb. 2005.

[30] Telecom Italia: Sanità Digitale. (2015). [Online]. Available: http://nuvolaitaliana.impresasemplice.it/digitalizzazione/sanit%C3%A0-digitale

[31] Qualcomm Life: 2net Platform. (2015). [Online]. Available: http://www.qualcommlife.com/wireless-health