

Quantitative Comparison of Human and Software Reliability in the Categorization of Sit-to-Stand Motion Pattern

Mirko Job^{ID}, Simone Battista^{ID}, Ronny Stanzani^{ID}, Alessio Signori^{ID}, and Marco Testa^{ID}

Abstract—The Sit-to-Stand (STS) test is used in clinical practice as an indicator of lower-limb functionality decline, especially for older adults. Due to its high variability, there is no standard approach for categorising the STS movement and recognising its motion pattern. This paper presents a comparative analysis between visual assessments and an automated-software for the categorisation of STS, relying on registrations from a force plate. 5 participants (30 ± 6 years) took part in 2 different sessions of visual inspections on 200 STS movements under self-paced and controlled speed conditions. Assessors were asked to identify three specific STS events from the Ground Reaction Force, simultaneously with the software analysis: the start of the trunk movement (Initiation), the beginning of the stable upright stance (Standing) and the sitting movement (Sitting). The absolute agreement between the repeated raters' assessments as well as between the raters' and software's assessment in the first trial, were considered as indexes of human and software performance, respectively. No statistical differences between methods were found for the identification of the Initiation and the Sitting events at self-paced speed and for only the Sitting event at controlled speed. The estimated significant values of maximum discrepancy between visual and automated assessments were 0.200 [0.039; 0.361] s in unconstrained conditions and 0.340 [0.014; 0.666] s for standardised movements. The software assessments displayed an overall good agreement against visual evaluations of the Ground Reaction Force, relying, at the same time, on objective measures.

Index Terms—Sensors activity recognition, kinematics, sensor systems.

I. INTRODUCTION

PERFORMANCE-BASED tests are important clinical tools used to identify functional decline in older adults and patients with either neurological or motor impairments [1]. Among them, the Sit-to-Stand test (STS) is strictly correlated with functional capacity and with independence in the

activities of daily living [2]–[9]. Furthermore, its simplicity allows it to be performed both in clinical and in-home environments across a broad range of patients' functional conditions [1], [5]. For these reasons, the STS movement is widely used in clinical research and practice, either as a single test [10]–[12] or as part of evaluation scales like the Short Physical Performance Battery [13]. Nonetheless, diversely from gait, a standardised categorisation for the STS motion pattern has yet to be established, due to its intrinsic variability. Furthermore, the lack of consensus in the definitions and measurement methodologies increases the uncertainty even in the definition of those events that normally are generally recognisable. For instance, the lifting of the thighs from the chair is defined across the literature as the peak instant of the horizontal [14] and vertical components [15] of the Ground Reaction Force (GRF), the time of maximum anterior head movement [16] or through a sensor-equipped chair [17].

In an attempt to describe a general model of the movement, Nuzik [18] divided the STS into two phases through the analysis of camera recordings, i.e. the flexion phase, consisting of the forward flexion of trunk and hip, and the extension phase, characterised by the lifting from the chair to the extension of knees, hips, and ankles. In later years, a finer sub-categorisation was introduced by Schenkman and colleagues [19]. They identified other two phases, with the use of a motion capture system, i.e. a “momentum transfer phase” between the flexion phase and the extension phase, when the inertia is transferred from the trunk flexion to the upper body, and a “stabilisation phase” when the hip is completely extended, and all the movements associated with stabilisation from rising are completed. Differently, Etnyre and colleagues analysed the kinetics of the STS under several conditions to recognise eleven invariably occurring events in the GRF, i.e. six in the vertical direction, three in the fore-aft direction, and two in the medial-lateral direction [20].

In general, despite the large amount of data obtained from a wide variety of sensors and technologies, it is still not possible to identify a univocal approach for a standardised definition of the STS motion pattern. Kinematic parameters proved to be strongly affected by the high within-between individuals' variability, limiting the performance of automated techniques to the recognition of dynamic transitions and static positions. On the contrary, kinetic variables permit a finer discretisation of the movement, but they still depend too much on visual evaluations to obtain reliable results [20], [21].

Manuscript received October 19, 2020; revised March 25, 2021; accepted April 11, 2021. Date of publication April 15, 2021; date of current version April 28, 2021. The work of Ronny Stanzani was supported by the Liguria Region European Social Fund Operational Program (POR-FSE) 2014–2020. (Corresponding author: Mirko Job.)

Mirko Job, Simone Battista, Ronny Stanzani, and Marco Testa are with the Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DiNOGMI), University of Genova, 16126 Genova, Italy, and also with the Rehabilitation and Engineering Laboratory (REHELab), Campus of Savona, University of Genova, 16126 Genova, Italy (e-mail: mirko.job@edu.unige.it).

Alessio Signori is with the Biostatistics Unit, Department of Health Sciences (DISSAL), University of Genova, 16126 Genova, Italy.

Digital Object Identifier 10.1109/TNSRE.2021.3073456

TABLE I
DESCRIPTIVE CHARACTERISTICS OF THE HAR DATASETS

HAR1		HAR2
19	N ^o participants	20
26±3	Age [years]	27±4
[9/10]	Gender [M/F]	[11/9]
63±10	Weight [kg]	67±15
171±9	Height [cm]	172±9

According to the previous statements, to establish an objective method for describing commonly occurring events in STS it is essential to assure quantitative and repetitive measures to be used as a reference in clinical practice and research.

II. METHODS

Objective: This study aims at evaluating the discrepancies between the performance of human assessment as well as a custom automated algorithm in the recognition of specific events of the STS movement identified from GRF profiles. The routine implementation is described in detail in an additional document (see Additional File 1).

A. Definition of the Datasets

We evaluated the performance of the proposed method on two Human Activity Recognition datasets (HAR1 and HAR2), collected at the REHELab (University Campus of Savona, Via Magliotto 2, 17100, Savona, Italy). Both HAR1 and HAR2 represent a series of sequential movements collected from a convenience sample of healthy young adults (Table I). The inclusion criteria for eligible participants were: good health, absence of musculoskeletal or neurological disorders and, ability to easily rise from a chair. Each participant had to sign an informed consent.

Participants performed the STS repetitions on a force plate (Kistler Winterthur, Switzerland). During the execution of the movement, a custom-made chair equipped with an electronic switch was able to record the time instants of rising (Seat Off) and sitting (Seat On). The aim of using a custom-made chair was to standardise the participants' position in each performance. The height of the chair was, therefore, regulating according to participants' height to reach a 90° knee flexion while sitting. In both datasets, participants had to execute two different tasks:

- To perform 10 repetitions of a single STS transition at self-paced speed (SP);
- To perform 10 repetitions of a single STS transition at a controlled speed (CT) with duration marked by a repetitive 4-second acoustic feedback, composed by a succession of 3 tones and a pause.

Between each repetition, one-minute break was necessary to reboot the STS software and to avoid participants from getting tired. For the same reasons, and to explain the execution of CT trial, five-minutes break was necessary between the SP and the CT trials. The different strategies of motor control involved in the two tasks outlined two typical profiles of the Ground Reaction Force (GRF). According to the definition of STS events given by Etnyre and colleagues [20] in SP trials, an initial deflection from the baseline was observed (Initiation). After reaching the lowest level in the force recording (Peak-counter),

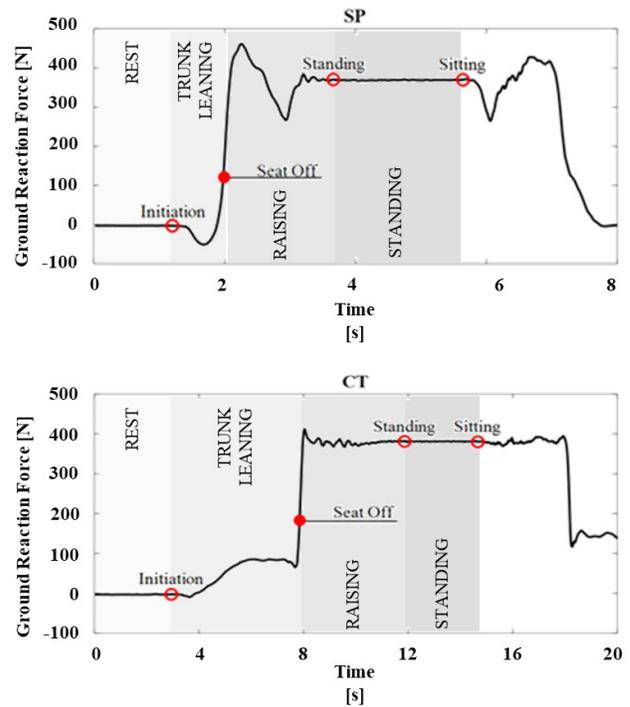


Fig. 1. GRF profiles of SP and CT trials with identification of STS events and phases.

the GRF raise to a global maximum (Peak) and subsequently levelled to a normal postural sway (Standing). Diversely, CT trials were characterised by a more gradual increase in the GRF following the progressive inclination of the trunk and the raising movement from the chair. Examples of force profiles for both SP and CT trials are displayed in Fig. 1.

Hence, the STS movement pattern was categorised in 4 sequential phases (Fig. 1):

- the **Resting** phase, identified as the initial sitting position;
- the **Trunk Leaning** phase, starting from the Initiation event to the Seat Off instant;
- the **Raising** phase, delimited from the Seat Off to the Standing event as the first instant of steady postural sway;
- the **Standing** phase, characterised by a stable upright position until the beginning of the sitting movement.

The two datasets are the result of two consecutive studies. Hence, some minor modification in the execution protocol of the exercise must be disclosed and taken into account. These changes are due to the progressive methodological improvement of the general work, which relies on previous empirical evidence to reduce possible future bias. The summary of the protocol differences between the two datasets is reported in Table II.

B. Evaluation of the Software Performance

In the lack of a proper gold standard approach, visual inspection remains the most reliable method to recognise STS events and phases on the base of GRF values [20]. Hence, simultaneously with the software analysis, five participants aged (30 ± 6) years visually identified the beginning of the trunk movement (Initiation event) and the limits of the stable stance (Standing and Sitting events) from the force

TABLE II
PROTOCOL DIFFERENCES BETWEEN HAR1 AND HAR2 DATASETS

HAR1	HAR2
In SP trials some participants tended to start the movement too early not allowing the registration of an appropriate resting phase.	In SP trials participants had to wait 3 seconds from the start of the acquisition before starting the movement.
In SP trials some participants tended to sit-down without having reached a sufficient stable standing phase.	In SP trials participants had to wait 3 seconds in standing phase to reach balance stability.
In CT trials participants considered the Stand-to-Sit transition as a single returning phase, starting from the beginning of the Sitting event until reaching the resting phase.	In CT trials participants were asked to control the descending movement, identifying two returning phases: a “Sitting” phase, starting from the beginning of the sitting movement until registration of the “seat-on” signal; a “Trunk Raising phase, where participants raise the trunk until reaching the resting phase. This was done in anticipation of future efforts in categorising the Stand-to-Sit transition.

TABLE III
RATERS’ DESCRIPTIONS

Age [years]	Professional background	Academic Level
29	Bioengineer	MSc/PhD student
27	Physicist	MSc/PhD student
40	Physiotherapist	PhD
25	Bioengineer	MSc/Research Fellow
30	Psychologist	MSc/PhD student

profiles of 200 STS sequences. The ages of the assessors are reported in Table II together with their professional and academic background.

As a measure of reliability, assessments were repeated in 2 distinct sessions, separated in time by a minimum interval of 1 hour. STS sequences were drawn randomly from a total pool of 742 acquisitions (HAR1 + HAR2) and were maintained across the measurement trials. This information was not made explicit to avoid possible learning effects. Among the 200 pooled sequences, 100 referred to SP trials, and 100 referred to CT trials. Before the first session, raters were briefly trained to recognise the onset of each event on five force profiles accordingly to the definitions given by Etnyre and colleagues [20], and an explanatory summary was always available in the form of MATLAB live script throughout all the measurements. All assessors were physiotherapists, bioengineers and PhD candidates with clinical experience in physiotherapy and expertise in movement analysis.

C. Data Analysis

Considering the possible sources of variability across different datasets, the similarity between movements was assessed using a correlation-based method, by comparing all the registered STS sequences with each other. This was done to evaluate the possible homogeneity of results obtainable from a combined analysis of the HAR1 and HAR2 datasets. For each pair of trials:

- through a cross-correlation operation, the relative lags that maximised the similarity between the two GRF profiles were found;
- the Pearson’s Correlation Coefficient (ρ) between the two STS trials, was calculated after shifting the respective GRF profiles according to the lags found in the first step.

This process was carried out separately for the SP and the CT trials. Distributions of ρ were extrapolated for each dataset comparison (i.e., HAR1 vs HAR2, HAR1 vs HAR1, HAR2 vs HAR2) and descriptive statistics were reported in terms of median and 95% Confidence Intervals (CI). Both the absolute

agreement between the repeated raters’ assessments as well as between the raters’ and the software’s assessment in the first trial, were evaluated through a non-parametric version of Bland-Altman statistics [22]. The former and the latter were considered as a quality index of the human’s and the software’s measuring performance, respectively. Normal distributions were tested using the Kolmogorov-Smirnov test and, furtherly investigated using the skewness and kurtosis indexes. To correct the analysis for possible outliers not directly connected to the evaluation of the raters or the performance of the software, data observations that fell outside 1.5 interquartile ranges above the upper quartile (75-th percentile) or below the lower quartile (25-th percentile), were visually inspected. In particular, wrong identifications resulting from an erroneous right mouse click, early movements or a malfunction of the electronic switch were eliminated. We calculated the systematic bias and the Limits of Agreement as the median of the absolute differences between assessments and the respective 2.5-th and 97.5-th percentile scores. The upper Limits of Agreement (ULoA) represented the maximum estimated error between the measures. The CI of the above-specified parameters were also calculated using a percentile bootstrap method based on 10k samples [23]. The percentile method was chosen for its conservative nature, as it tends to produce wider CI less sensitive to population value and sample size [24]. To compare the maximum errors made between raters’ and human-software’s agreement, we used a two-tailed two-sample t-test [25], exploring the significant differences between measures. Every ULoAs and respective CI were approximated to normal distributions characterised by mean values μ_i and standard deviation σ_i [26] calculated in Eq. 1 and Eq. 2.

The choice of the lower or upper confidence limit is guided by the need to calculate the largest standard error to obtain a more conservative approximation. The statistical analysis was stratified by considering the different STS events separately, dividing the results obtained in SP and CT trials. Bland-Altman statistics was implemented in MATLAB as a modified version of the *BlandAltman.m* function developed by Ran Klein from the Department of Nuclear Medicine of the Ottawa Hospital [27]. The two-tailed two-sample t-test was executed with the online “Comparison of means calculator” tool from Medcalc Statistical Software [28].

III. RESULTS

The results of the correlation-based analysis, for testing the homogeneity of the combined dataset, are displayed in

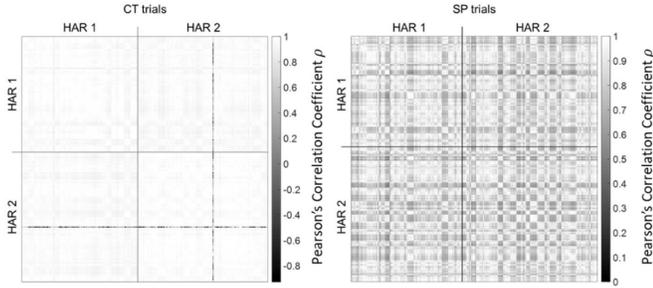


Fig. 2. Correlograms of the ρ coefficients for each pairwise comparison between the trials in HAR1 and HAR2.

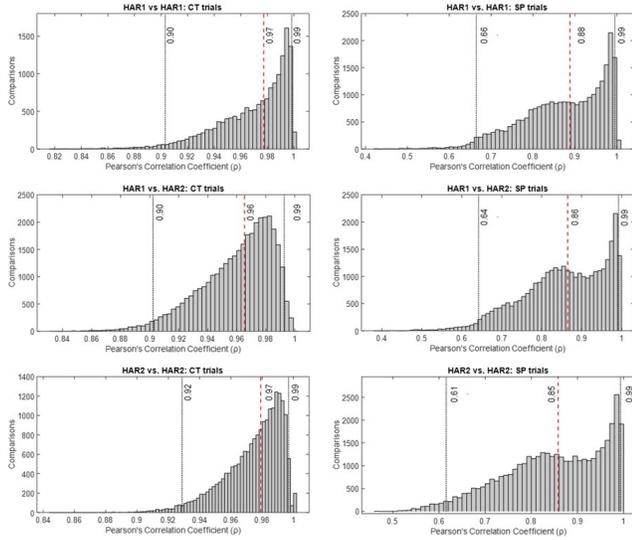


Fig. 3. Histograms of the Pearson's Correlation Coefficients in within and between datasets comparison. The median is indicated by the dashed red line and the CI are indicated by the dotted black line.

the correlograms in Fig 2 and the descriptive distributions (medians and CI) of ρ are illustrated in the histograms in Fig 3.

In general, the correlograms highlighted strong homogeneity across the combined dataset, with high correlations ($\cong 1$) across all CT trials comparing STS sequences from the same (HAR1 vs HAR1, HAR2 vs HAR2) or from different dataset (HAR1 vs HAR2) and sequences from the same or different participants. In SP trials, the correlations were moderate to strong, significantly lower than CT trials. However, the descriptive distributions of ρ in the within-between comparisons for both CT and SP trials were comparable both in terms of descriptive statistics and shape.

The Bland-Altman plots relating to the identification of the Initiation, Standing and Sitting events are shown respectively

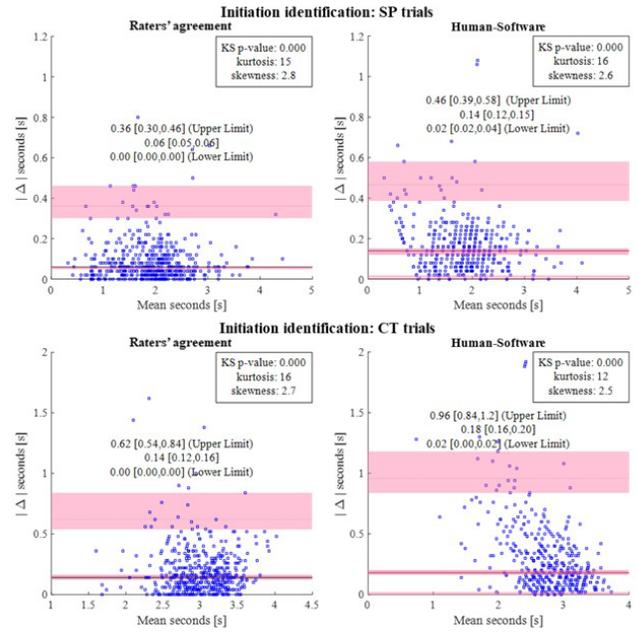


Fig. 4. Bland-Altman plots depicting the raters' agreement (left) and the human-software's agreement (right) of the identification of the Initiation event in SP trials (upper plots) and CT trials (lower plots).

in Fig. 4, Fig. 5 and Fig. 6 and the summary of the descriptive statistics of bias and the ULoAs (CI) are reported in Table IV.

In both SP and CT trials, for all STS events, participants showed significantly lower identification bias in the repeated assessments, compared to the difference displayed against the measurements performed by the software. Moreover, the raters' agreement in the identification of the STS events significantly decreases in CT trials with a consequent increase in the value of the systematic error.

The results of the two-tailed two-sample t-test are summarised in Table V.

Significant differences in ULoAs between raters' evaluation and Human-Software assessments were found in the Standing event identification for both SP (0.200s [0.039; 0.361], $p > 0.05$) and CT (0.300s [0.017; 0.583], $p > 0.05$) trials. A significant difference ($p > 0.05$) of 0.340s [0.014; 0.666] was also found in the Initiation event identification in CT trials.

IV. DISCUSSION

This work aimed at evaluating the performance of a new automated approach for the recognition of clinically relevant events in the STS and at comparing its performance to the human visual assessment. The results obtained offer a double contribution in prospective researches since not only we quantified the discrepancy between the two methods, but we also compared it against the maximum error made in repeated

$$\mu_i = ULoA_i \quad (1)$$

$$\sigma_i = \begin{cases} \frac{|CI_{up} - (ULoA)_i|}{1.96} \cdot \sqrt{N} & \text{if } |CI_{up} - (ULoA)_i| > |CI_{low} - (ULoA)_i| \\ \frac{|CI_{low} - (ULoA)_i|}{1.96} \cdot \sqrt{N} & \text{if } |CI_{low} - (ULoA)_i| > |CI_{up} - (ULoA)_i| \end{cases} \quad (2)$$

TABLE IV
DESCRIPTIVE STATISTICS FOR BIAS AND ULOA FOR EVERY COMPARISON. (*) $P < 0.05$ FOR NOT OVERLAPPING CI

Events	SP trials						CT trials					
	Raters' Agreement			Human-Software			Raters' Agreement			Human-Software		
	Bias [s]	ULoA [s]	N									
Initiation	0.06 [0.05-0.06]	0.36 [0.30-0.46]	492	0.14 [0.12-0.15]	0.46 [0.39-0.58]	490	0.14 [0.12-0.16]	0.62 [0.54-0.84]	483	0.18 [0.16-0.20]	0.96 [0.84-1.20]	493
Standing	0.10 [0.08-0.10]	0.62 [0.48-0.74]	488	0.30 [0.28-0.32]	0.82 [0.74-0.86]	495	0.16 [0.16-0.20]	1.30 [1.10-1.70]	484	0.44 [0.40-0.48]	1.60 [1.40-1.70]	493
Sitting	0.08 [0.06-0.08]	0.42 [0.34-0.47]	497	0.20 [0.18-0.22]	0.48 [0.46-0.60]	494	0.14 [0.12-0.14]	0.69 [0.60-0.86]	491	0.20 [0.18-0.22]	0.81 [0.66-0.86]	490

TABLE V
TWO-TAILED TWO-SAMPLE T-TEST RESULTS. DIFFERENCES WITH P-VALUES < 0.05 WERE CONSIDERED STATISTICALLY SIGNIFICANT (*)

	SP trials			CT trials		
	Initiation	Standing	Sitting	Initiation	Standing	Sitting
Difference [s]	0.100	0.200	0.060	0.340	0.300	0.120
95% CI [s]	-0.056	0.039	-0.084	0.014	0.017	-0.107
	0.256	0.361	0.204	0.666	0.583	0.347
t-statistic	1.255	2.440	0.816	2.045	2.079	1.037
df	980	981	989	974	975	979
p-value	0.209	0.015 (*)	0.414	0.041 (*)	0.038 (*)	0.300

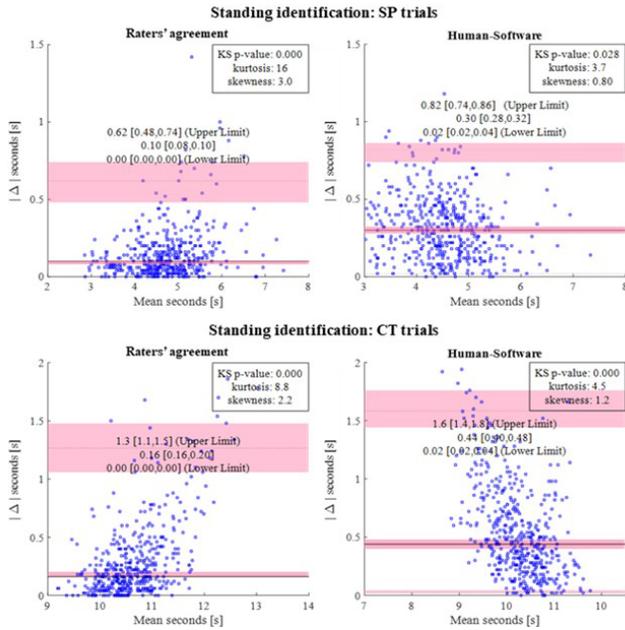


Fig. 5. Bland-Altman plots depicting the raters' agreement (left) and the human-software's agreement (right) of the identification of the Standing event in Task A (upper plots) and Task B (lower plots).

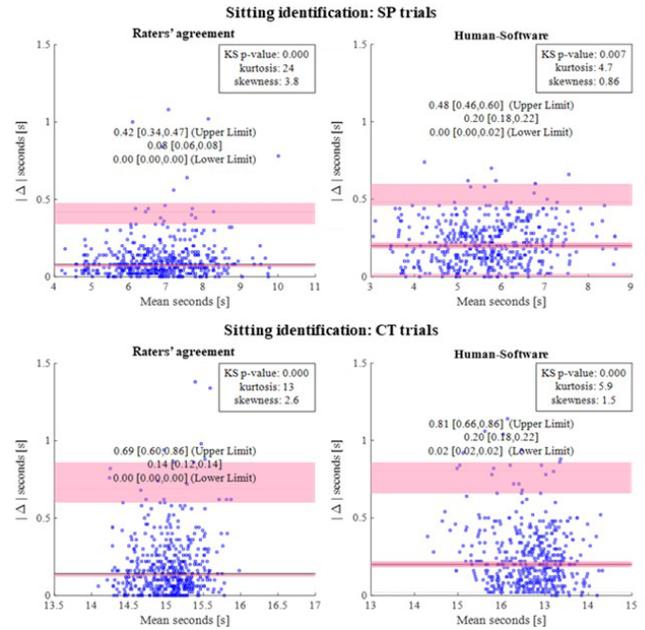


Fig. 6. Bland-Altman plots depicting the raters' agreement (left) and the human-software's agreement (right) of the identification of the Standing event in SP trials (upper plots) and CT trials (lower plots).

visual measures. Despite the significant lower systematic bias in repeated evaluations, the comparison between visual assessments and the proposed approach showed similar values of maximum absolute error.

More specifically no statistical differences were found in the identification of the Initiation and the Sitting events in SP trials and in the identification of only the Sitting event in CT trials. The worsening of the observed agreement during CT movements was generally in line with our expectation, as ULoAs values could be affected by uncertainties due to the kinetic modifications resulting from the standardisation of the movement. As an example, the initial GRF's signal deflection effect has been related to the contraction of the hip flexor muscles accompanied by an early lifting of the thighs and a short discharge on the load of the feet [46]. This

effect is highly dependent on each individual's movement strategy [20]. Previous studies on the STS movement in both healthy younger people and older adults outlined two main movement strategies [18], [19], [47]–[51]. The first is typically observed in younger people since it implies high balance control and greater momentum on the lower limbs. The second is typically used by the elderly, who firstly perform an accentuated flexion of the trunk toward the knees, moving the centre of mass over the feet. This strategy is slower but presents a reduction of global instability. When the person is forced to get up in a controlled way, the strategy used resembles the elderly's one. This causes two effects on the deflection phenomenon on the GRF signal. The first is that its amplitude becomes smaller, since the lifting is delayed after the trunk flexion and consequently the hip muscles contraction

is reduced. The second is that the discharges phases on the GRF become two: the first when the person begins to lean forward, the second when the desired inclination has been reached and the thrust preceding the lift-off occurs. These two effects contribute to the reduction in the detectability of the Initiation event, which appears less pronounced in the controlled lift.

Another important consideration highlights the intrinsic subjectiveness of human evaluations [29]. One could consider the slightest oscillation either as an extension of a contiguous static phase or as the limit of a movement transition. Moreover, visual assessments can vary across repeated measures and differ in individuals, depending on their professional experience [30], [31].

Nonetheless, with a maximum estimated discrepancy of 0.200 s [0.039; 0.361] and 0.340 s [0.014; 0.666] respectively for normal and standardised speed, the analysed routine may not be suitable for evaluations on a single patient, where the expertise of health professionals plays a key role in the diagnosis process, often requiring a high level of abstraction [29]. However, the capacity of automatically and objectively handling big quantity of data, would free health professionals from the encumbrance of the processing, with a reasonable margin of error. Moreover, it is also true that a great quantity of data collected through sensor systems is inherently noisy, and for this reason, its analysis should consider and handle some degree of uncertainty [32]. In this context, the presented algorithm can be used as an objective labelling procedure for artificial intelligence methods to provide reliable ground truth data for supervised classification tasks in the field of big data analytics and Human Activity Recognition [33], [34].

This is the first study that aims at comparing human and automated assessments in the identification of the events that divide the different phases of the STS movement, relying on data collected through a force plate. Previous works [35]–[39] evaluated the performances of various algorithms developed for the identification of Sit-to-Stand and Stand-to-Sit postural transitions using data acquired from inertial sensors. A recent paper of Atrsaei and colleagues [40] validated the accuracy of a new routine based on a single device against visual assessments on-camera recordings of STS movements, obtaining levels of agreement above 94%, in terms of positive predictive values and sensitivity. As a direct comparison with the present study, the use of inertial sensors is usually preferable since they can be also applied in non-clinical environments [41]. However, their measurements are strongly influenced by the inter- and intra-individual variability of the movement [20], [42], limiting the recognition of the STS motion pattern to the simple discrimination of static and dynamic phases. Conversely, our choice to use a force plate has some doubtless limitations in terms of costs and portability but the strong advantage of providing easier interpretable results, on which it is possible to identify clinically significant movement events and phases. Under this point of view, supported by the recent advances in the field of machine learning, it is possible to consider this data as a valid ground-truth reference to train specific IMU-based approaches in a finer recognition of the STS. This

transition could allow the development of accessible, wearable rehabilitation tools, combining the discriminating power of gold-standard instruments with the limited dimensions and costs of inertial sensors.

The results highlighted in this work should be appraised in consideration of some limitations. The population chosen to perform the STS was young and did not suffer from any musculoskeletal and neurological conditions. This allowed for evaluating the difference in the discrepancies of the STS performance assessment in a population whose movement patterns are not hindered by the abovementioned conditions. Similarly, also the standardisation of posture at the beginning of the trials could have restricted the applicability of the presented evidence. Both the setting of the starting position and the individuals' characteristics influence the STS motion strategy and consequently, the value of the vertical GRF and the successful outcome of the movement [20], [43], [44]. Hence, future studies should investigate if the presence of different conditions (e.g., different uses of the arms during the rising, different fatigue condition of the participants etc.) and different population groups might widen the discrepancies in the assessment of STS performance. Moreover, the limited sample size and the heterogeneous professional background of the raters could have impacted the final estimate of the visual measurements. Further evaluations performed by a larger sample of qualified health professionals could provide a more precise picture of the reliability of the proposed method.

V. CONCLUSION

The presented algorithm displayed an overall good agreement against visual assessments in the identification of significant STS events from values of vertical Ground Reaction Force. By providing objective and fast measures, this automated method can be applied for the analysis of a large amount of data to identify specific trends and patterns in patients' movement. Further studies should be undertaken to validate its accuracy across different types of patients' conditions.

ACKNOWLEDGMENT

This work was developed within the framework of the DINOEMI Department of Excellence of MIUR 2018-2022 (law 232/2016). This work was conducted on two pre-existing and anonymous databases (HAR1 and HAR2) on young and healthy subjects; therefore, the Ethical Committee approval was not required.

REFERENCES

- [1] R. Soubra, A. Chkeir, and J.-L. Novella, "A systematic review of thirty-one assessment tests to evaluate mobility in older adults," *Bio-Med Res. Int.*, vol. 2019, pp. 1–17, Jun. 2019. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6610744/>
- [2] R. W. Bohannon, "Measurement of sit-to-stand among older adults," *Topics Geriatric Rehabil.*, vol. 28, no. 1, pp. 11–16, Jan. 2012.
- [3] P. M. Dall and A. Kerr, "Frequency of the sit to stand task: An observational study of free-living adults," *Appl. Ergonom.*, vol. 41, no. 1, pp. 58–61, Jan. 2010.
- [4] M. M. Gross, P. J. Stevenson, S. L. Charette, G. Pyka, and R. Marcus, "Effect of muscle strength and movement speed on the biomechanics of rising from a chair in healthy elderly and young women," *Gait Posture*, vol. 8, no. 3, pp. 175–185, Dec. 1998.
- [5] S. Hellmers *et al.*, "Measurement of the chair rise performance of older people based on force plates and IMUs," *Sensors*, vol. 19, no. 6, p. 1370, Mar. 2019.

- [6] S. Hellmers *et al.*, "Towards a minimized unsupervised technical assessment of physical performance in domestic environments," in *Proc. 11th EAI Int. Conf. Pervas. Comput. Technol. Healthcare*, New York, NY, USA: ACM, May 2017, pp. 207–216, doi: [10.1145/3154862.3154882](https://doi.org/10.1145/3154862.3154882).
- [7] L. M. Inkster and J. J. Eng, "Postural control during a sit-to-stand task in individuals with mild Parkinson's disease," *Exp. Brain Res.*, vol. 154, no. 1, pp. 33–38, Jan. 2004.
- [8] J. Jeyasurya, H. F. M. Van der Loos, A. Hodgson, and E. A. Croft, "Comparison of seat, waist, and arm sit-to-stand assistance modalities in elderly population," *J. Rehabil. Res. Develop.*, vol. 50, no. 6, pp. 835–844, 2013.
- [9] J. S. Munton, M. I. Ellis, M. A. Chamberlain, and V. Wright, "An investigation into the problems of easy chairs used by the arthritic and the elderly," *Rheumatology*, vol. 20, no. 3, pp. 164–173, 1981.
- [10] R. W. Bohannon, "Reference values for the five-repetition sit-to-stand test: A descriptive meta-analysis of data from elders," *Perceptual Motor Skills*, vol. 103, no. 1, pp. 215–222, Aug. 2006.
- [11] M. Csuka and D. J. McCarty, "Simple method for measurement of lower extremity muscle strength," *Amer. J. Med.*, vol. 78, pp. 77–81, Dec. 1985.
- [12] C. J. Jones, R. E. Rikli, and W. C. Beam, "A 30-S chair-stand test as a measure of lower body strength in community-residing older adults," *Res. Quart. Exerc. Sport*, vol. 70, no. 2, pp. 113–119, Jun. 1999.
- [13] J. M. Guralnik *et al.*, "A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission," *J. Gerontol.*, vol. 49, no. 2, pp. M85–M94, Mar. 1994.
- [14] C. A. M. Doorenbosch, J. Harlaar, M. E. Roebroeck, and G. J. Lankhorst, "Two strategies of transferring from sit-to-stand; the activation of monoarticular and biarticular muscles," *J. Biomech.*, vol. 27, no. 11, pp. 1299–1307, Nov. 1994.
- [15] B. K. Kaya, D. E. Krebs, and P. O. Riley, "Dynamic stability in elders: Momentum control in locomotor ADL," *J. Gerontol. Ser. A, Biol. Sci. Med. Sci.*, vol. 53A, no. 2, pp. M126–M134, Mar. 1998.
- [16] N. B. Alexander, A. B. Schultz, and D. N. Warwick, "Rising from a chair: Effects of age and functional ability on performance biomechanics," *J. Gerontol.*, vol. 46, no. 3, pp. M91–M98, May 1991.
- [17] R. C. van Lummel, E. Ainsworth, J. M. Hausdorff, U. Lindemann, P. J. Beek, and J. H. van Dieën, "Validation of seat-off and seat-on in repeated sit-to-stand movements using a single-body-fixed sensor," *Physiol. Meas.*, vol. 33, no. 11, pp. 1855–1867, Nov. 2012.
- [18] S. Nuzik, R. Lamb, A. VanSant, and S. Hirt, "Sit-to-stand movement pattern," *Phys. Therapy*, vol. 66, no. 11, pp. 1708–1713, Nov. 1986.
- [19] M. Schenkman, R. A. Berger, P. O. Riley, R. W. Mann, and W. A. Hodge, "Whole-body movements during rising to standing from sitting," *Phys. Therapy*, vol. 70, no. 10, pp. 638–648, Oct. 1990.
- [20] B. Etnyre and D. Q. Thomas, "Event standardization of sit-to-stand movements," *Phys. Therapy*, vol. 87, no. 12, pp. 1651–1666, Dec. 2007.
- [21] L. Piano, T. Geri, and M. Testa, "Raising and stabilization phase of the sit-to-stand movement better discriminate healthy elderly adults from young subjects: A pilot cross-sectional study," *Arch. Physiotherapy*, vol. 10, no. 1, pp. 1–8, Dec. 2020. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7161000/>
- [22] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, pp. 307–310, Dec. 1986.
- [23] D. Kushary, "Bootstrap methods and their application," *Technometrics*, vol. 42, no. 2, pp. 216–217, May 2000.
- [24] K. Jung, J. Lee, V. Gupta, and G. Cho, *Comparison of Bootstrap Confidence Interval Methods for GSCA Using a Monte Carlo Simulation*. Accessed: Jul. 5, 2020. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6797821/>
- [25] D. G. Altman, *Practical Statistics for Medical Research*. Boca Raton, FL, USA: CRC Press, 1990.
- [26] *Choosing Effect Measures and Computing Estimates of Effect*. Accessed: Jul. 20, 2020. [Online]. Available: <https://training.cochrane.org/handbook/current/chapter-06>
- [27] *Bland-Altman and Correlation Plot*. Accessed: Jul. 5, 2020. [Online]. Available: <https://it.mathworks.com/matlabcentral/fileexchange/45049-bland-altman-and-correlation-plot>
- [28] F. S. MedCalc's. *Comparison of Mean Calculator*. Accessed: Jul. 5, 2020. [Online]. Available: https://www.medcalc.org/calc/comparison_of_means.php
- [29] A. Connor, *Clinical Instruction & Evaluation: A Teaching Resource*. Burlington, MA, USA: Jones & Bartlett, 2014.
- [30] T. Jaarsma, H. Jarodzka, M. Nap, J. J. G. van Merriënboer, and H. P. A. Boshuizen, "Expertise in clinical pathology: Combining the visual and cognitive perspective," *Adv. Health Sci. Educ.*, vol. 20, no. 4, pp. 1089–1106, Oct. 2015.
- [31] A. M. Persky and J. D. Robinson, "Moving from novice to expertise and its implications for instruction," *Amer. J. Pharmaceutical Educ.*, vol. 81, no. 9, p. 6065, Nov. 2017.
- [32] R. H. Hariri, E. M. Fredericks, and K. M. Bowers, "Uncertainty in big data analytics: Survey, opportunities, and challenges," *J. Big Data*, vol. 6, no. 1, Dec. 2019.
- [33] T. Li *et al.*, "Automatic timed Up-and-Go sub-task segmentation for Parkinson's disease patients using video-based activity classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2189–2199, Nov. 2018.
- [34] B. R. Greene, E. P. Doheny, R. A. Kenny, and B. Caulfield, "Classification of frailty and falls history using a combination of sensor-based mobility assessments," *Physiol. Meas.*, vol. 35, no. 10, pp. 2053–2066, Oct. 2014.
- [35] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 711–723, Jun. 2003.
- [36] A. Godfrey, A. K. Bourke, G. M. Ólaighin, P. van de Ven, and J. Nelson, "Activity classification using a single chest mounted tri-axial accelerometer," *Med. Eng. Phys.*, vol. 33, no. 9, pp. 1127–1135, Nov. 2011.
- [37] A. Salarian, H. Russmann, F. J. G. Vingerhoets, P. R. Burkhard, and K. Aminian, "Ambulatory monitoring of physical activities in patients with Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 12, pp. 2296–2299, Dec. 2007.
- [38] M. H. Pham *et al.*, "Wearable-based sit-to-stand and stand-to-sit algorithm for patients with parkinson's disease and older adults in a home-like environment," *Frontiers Neurol.*, vol. 9, p. 51, Aug. 2018. Available from: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00652/full>
- [39] D. Rodríguez-Martín, A. Samã, C. Pérez-López, and A. Catalã, "Identification of sit-to-stand and stand-to-sit transitions using a single inertial sensor," *Stud. Health Technol. Informat.*, vol. 177, pp. 113–117, Dec. 2012.
- [40] A. Atrsaei *et al.*, "Postural transitions detection and characterization in healthy and patient populations using a single waist sensor," *J. NeuroEng. Rehabil.*, vol. 17, no. 1, p. 70, Dec. 2020.
- [41] M. Job, A. Dottor, A. Viceconti, and M. Testa, "Ecological gait as a fall indicator in older adults: A systematic review," *Gerontologist*, vol. 60, no. 5, pp. e395–e412, Jul. 2020.
- [42] M. Kristiansen, G. H. F. Rasmussen, M. E. Sloth, and M. Voigt, "Inter- and intra-individual variability in the kinematics of the back squat," *Hum. Movement Sci.*, vol. 67, Oct. 2019, Art. no. 102510.
- [43] H.-J. Noh, C.-Y. Kim, H.-D. Kim, and S.-W. Kim, "Changes in muscle activation and ground reaction force of the lower limbs according to foot placement during sit-to-stand training in stroke patients," *Amer. J. Phys. Med. Rehabil.*, vol. 99, no. 4, pp. 330–337, Apr. 2020.
- [44] S. Shen, T. Abe, T. Tsuji, K. Fujii, J. Ma, and T. Okura, "The relationship between ground reaction force in sit-to-stand movement and lower extremity function in community-dwelling Japanese older adults using long-term care insurance services," *J. Phys. Ther. Sci.* vol. 29, pp. 15661–15669, Oct. 2017.
- [45] M. Job. (2020). *Sit-to-Stand Categorisation*. [Online]. Available: <http://doi.org/10.5281/zenodo.3956138>
- [46] J. H. Carr and A. M. Gentile, "The effect of arm movement on the biomechanics of standing up," *Hum. Movement Sci.*, vol. 13, no. 2, pp. 175–193, Jun. 1994.
- [47] M. A. Hughes, B. S. Myers, and M. L. Schenkman, "The role of strength in rising from a chair in the functionally impaired elderly," *J. Biomechan.*, vol. 29, no. 12, pp. 1509–1513, Dec. 1996, doi: [10.1016/S0021-9290\(96\)80001-7](https://doi.org/10.1016/S0021-9290(96)80001-7).
- [48] D. M. Scarborough, C. A. McGibbon, and D. E. Krebs, "Chair rise strategies in older adults with functional limitations," *J. Rehabil. Res. Develop.*, vol. 44, no. 1, p. 33, 2007, doi: [10.1682/jrrd.2005.08.0134](https://doi.org/10.1682/jrrd.2005.08.0134).
- [49] N. B. Alexander, D. J. Koester, and J. A. Grunawalt, "Chair design affects how older adults rise from a chair," *J. Amer. Geriatrics Soc.*, vol. 44, no. 4, pp. 356–362, Apr. 1996, doi: [10.1111/j.1532-5415.1996.tb06402.x](https://doi.org/10.1111/j.1532-5415.1996.tb06402.x).
- [50] E. Papa and A. Cappelozzo, "Sit-to-stand motor strategies investigated in able-bodied young and elderly subjects," *J. Biomechan.*, vol. 33, no. 9, pp. 1113–1122, Sep. 2000, doi: [10.1016/S0021-9290\(00\)00046-4](https://doi.org/10.1016/S0021-9290(00)00046-4).
- [51] S. DM, K. DE, and H. BA, "Quadriceps muscle strength and dynamic stability in elderly persons," *Gait Posture*, vol. 10, no. 1, pp. 10–20, Sep. 1999, doi: [10.1016/s0966-6362\(99\)00018-1](https://doi.org/10.1016/s0966-6362(99)00018-1).