Automatic Scoring of Chair Sit-to-Stand Test Using a Smartphone

Arshad Sher[1], David Langford[2], Federico Villagra[2], and Otar Akanyeti[1*]

[1] Department of Computer Science, Aberystwyth University, Aberystwyth, Ceredigion SY23 3FL, UK.
[2] Institute of Biology, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, Ceredigion SY23 3FL, UK.

`* corresponding author (ota1@aber.ac.uk)`

Abstract. Chair sit to stand tests (CST) are widely used in clinical settings to measure endurance, balance and lower extremity muscle strength. It also allows clinicians to predict falls and cognitive decline in older adults. The current CST measurements are done manually using a timer. The manual CST measurements can be imprecise (often leading to high inter-rater variability), and they ignore what kinematics strategies participants use to stand up and sit back on the chair. In this study, we present a smartphone based automatic CST analysis system. The system has the ability of generating a CST score, and perform cycle by cycle motion analysis. To achieve this, it employs two XGBoost classifiers one for recognising who is taking the test and which chair rising strategy they use. This information is then used to adapt its algorithms for more accurate CST score predictions. The performance of the system was tested on 30 participants including three demographics group (healthy young, healthy adult and Parkinson's) who were using two different chair rising strategies (flexion and momentum transfer). Overall, the system had above 95% classification accuracy, and the mean absolute difference between predicted and actual CST cycle completion time was less than 60 ms ($< 10\%$ considering that the average cycle completion time was above 1 second). These results are highly encouraging towards developing a new smartphone-based gait and balance assessment tool that can be used in outdoor settings.

Keywords: chair sit to stand, chair rising strategies, older adults, wearable sensors, smartphone, computational intelligence

1 Background

Successful transition from sitting to standing requires motor control and lower and upper body coordination. Most falls, reported in older adults or people suffering from chronic movement disorders, occur during sit to stand [1]. Chair sit to stand tests (CST) are widely used in clinical settings to measure lower extremity strength and endurance, and CST scores are often correlated with fitness, falls and cognitive decline [2]. During CSTs, participants repeatedly stand up and sit back on a chair until they are told to stop. The CST performance can be measured in two different ways: 1) by counting the number of successful repetitions in a fixed amount of time (e.g., 30 seconds) or 2) by measuring the amount of time required to reach a desired number of successful repetitions (e.g., 5 repetitions). More repetitions (first way) or less time (second way) indicate better health status.

Usually, clinicians use a stopwatch to track time and count the number of repetitions manually. The CST scores are evaluated together with other functional fitness tests such as 10-meter walk, timed-up-and-go, Tinetti balance scale or Berge balance scale for diagnosing motor symptoms and postural stability in older adults [3–6]. CSTs are often performed in laboratory settings, and participants are given very specific instructions on how to complete the test (such as "hands on chest", "sit still", "lean back", "bend forward", "push hips and knees to fully extend or push up", "flex knees" and "sit back"). However, there is no guarantee that participants would perform the test correctly resulting in undesired variations in assessment.

In addition, how people rise from a chair varies across the populations of young, old and disabled individuals. There are two common chair rising strategies (CRS) during CSTs: 1) flexion strategy where exaggerated forward leaning occurs before rising (balancing upper body over knees) (Fig. 1-a). In this strategy, vertical movements are less evident; the body does not reach to the fully extended complete standing posture [7, 8], and 2) momentum transfer (MT) strategy where forward leaning and rising from the chair occur almost simultaneously followed by full hip, back and knees extensions to complete standing (Fig. 1-b). This strategy is typically adopted by physically active individuals with strong legs. Frail older adults or people with chronic impairments are likely to use flexion strategy due to weak legs and fear of falling [1]. However, the CRS choice may also depend on other factors including fatigue, age, injury or cognitive load [9, 10].

Manual CST measurements are time consuming, imprecise (often leads to high inter-rater variability) and do not quantify how people are coordinating lower and upper body parts to achieve postural stability. In recent years, wearable devices housing inertial measurement unit (IMU) sensors (custom built or commercial such as smartphones) have been used to automate CST assessments measuring clinically relevant parameters [2]. Cerrito et al. used a smartphone to estimate CST cycle completion time in older adults. The phone predictions were compared to actual measurements recorded using a force plate. The mean absolute difference was 90 ± 130 milliseconds [11]. Cobo $etal.$ used a smartphone to count the number of repetitions during 30-second CST in older adults [2]. The phone was placed around the thigh, CST cycles were detected using changes in thigh angle. The average prediction error was two cycles. The existing studies mostly focused on estimating CST scores without investigating CST kinematics including CRS, which is the main focus of this study.

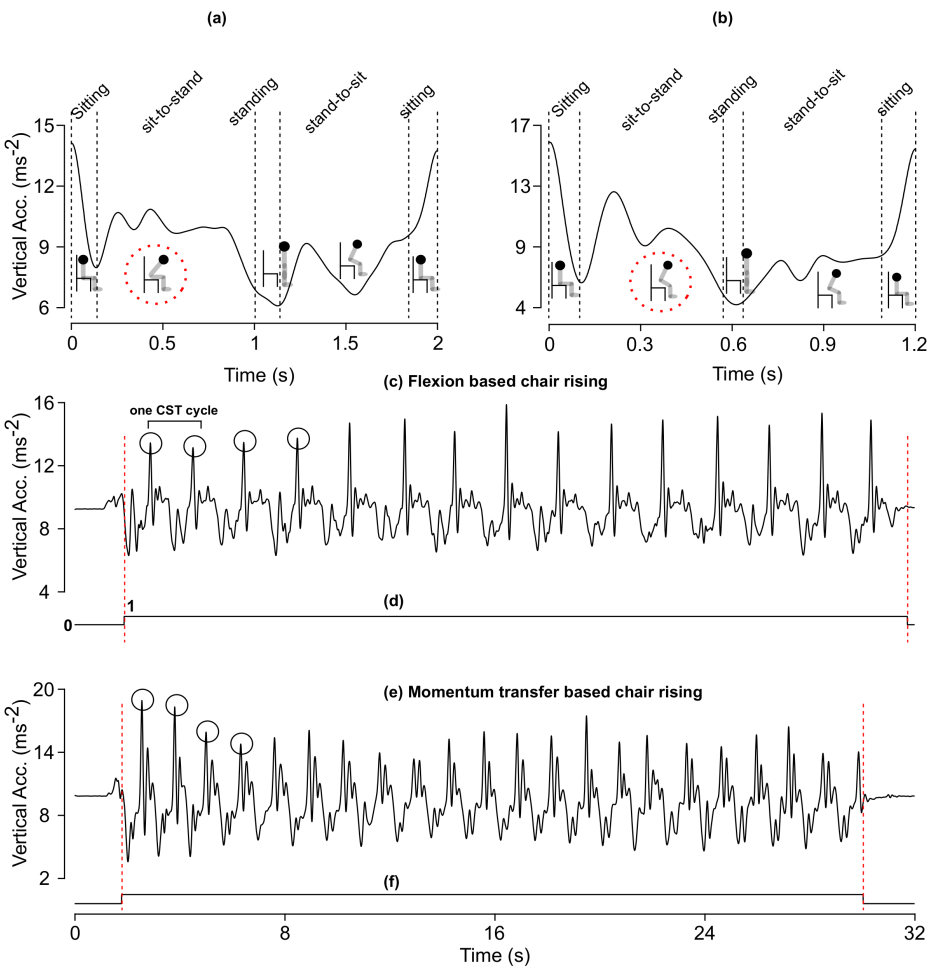

Fig. 1: Representation of CRS with accelerometer data: **(a)** shows one CST cycle for flexion CRS. **(b)** a complete cycle from MT CRS. **(c)** Complete CST of flexion CRS with more time between adjacent peaks. **(d, f)** is the presentation of binary signal which used to detect start and end of the CST. **(e)** is the MT strategy which clearly shows higher number of cycles than flexion CRS. **(c, e)** empty circles indicate the start and end of the CST cycle.

We present a novel software system for automatic end-to-end CST analysis using smartphones. The system is capable of: 1) segmenting CST events, 2) distinguishing between participants with gait impairments and healthy controls, 3) recognising CRS, 4) generating CST score, and 5) performing cycle by cycle motion analysis. The long-term goal is to create a standalone smartphone technology that can be used for objective gait and balance assessments in clinical and community settings.

The paper is organised as follows. Section 2 starts with the description of data collection, protocols for participant selection and the experimental design, as well as the implementation of the proposed software system. The experimental results are given in Section 3. Section 4 highlights the main findings and compares results with those reported in the literature. Section 5 concludes the paper with a short summary and future work.

2 Materials and methods

2.1 Data collection

Data recording mobile app A custom-built android app was used to record from a smartphone's IMU sensors. Prior to data collection, the app prompts the user to enter the participant's demographic details (Table 1) and expected duration of the test. After the test, the recorded data is uploaded to an online secure server from where it is downloaded for off-line processing and analysis.

Table 1: Demographic information of the participants.

Different categories of participants			
Parameters	Healthy young	Healthy old	PD
Age (years)	43.5 ± 10.6	67.5 ± 8.2	70.3 ± 11.06
Weight (kg)	66.7 ± 14.23	78.7 ± 10.6	68.7 ± 13.63
Height (cm)	173 ± 9.03	177.6 ± 6.7	167.6 ± 9.6
BMI (kg m^{-2})	22.3 ± 1.5	24.9 ± 1.6	24.6 ± 1.4
Females	2	8	1
Males	3	12	4
Flexion	2	8	4
MT	3	12	1

Subjects 30 participants voluntarily took part in the study (signing the informed consent form). Participants were comprised of: i) Five healthy young adults ($<$ 60 years old), ii) 20 healthy old adults ($>$ 60 years old), and iii) five adults with Parkinson's (no age limit), all recruited from Ceredigion county (mid Wales, UK). The criteria to participate in the study was simple, if a participant can stand without support and follow the guidelines of the test, they were allowed to take part in the study. This research was approved by an independent board of Ethics Committee at the Aberystwyth University, and all experimental protocols were inline with the Declaration of Helsinki guidelines.

Experiments All experiments were carried out in the human biomechanics laboratory at Aberystwyth University over the summer of 2018. A fixation belt with a phone holder was tied around the waist without discomforting the participants. The phone (Google Pixel II, sampling rate = 400 Hz) was placed in the phone holder horizontally in such a way that the accelerometer x-axis was recording vertical movements, y-axis recording side movements and z-axis recording forward movements of the body.

A firm armless chair with a straight back was used for all the tests. Prior to actual recordings, participants were shown how to perform the test and were asked to practice few times to familiarise with the test. All participants were instructed to perform a 30-second CST as: 1) sit in the centre of the chair with straight back as much as possible (with Parkinson's and old subjects sitting with straight back was not always possible), 2) feet should be shoulder width apart and positioned on the floor such that knees are slightly behind the ankles making angle greater than 90 and less than 115 degrees, 3) both hands should be crossed around wrists and placed on the chest.

The actual CST recordings were performed as follows: the experimenter started the camera recording, started the phone recording, placed the phone in the phone holder while participants were sitting on the chair, asked participants if they were ready to start the test, waited for few seconds after receiving participants confirmation, gave the start command and simultaneously started the stop watch, gave the stop command when 30 seconds was up, stopped the watch, waited for few seconds, took out the phone from the holder, stopped the phone recording, and stopped the camera recording.

All participants performed CST at their comfortable speed to avoid falls, and they were allowed to stop at any time. The tests were carried out under the supervision of Mr. David Langford (co-author) who is a registered clinical exercise physiologist. For ground truth, all experiments were recorded using a GoPro 6 camera from a side view (frame rate = 60 frames s^{-1}). These videos were annotated manually to select participant group (young, old or PD), label CRS (flexion or MT), detect sit-to-stand as well as stand to sit cycles, and count the number of successful repetitions. The participant info was also kept in the meta data.

2.2 Software system

All phone recordings were analyzed using the new software system which is comprised of five data processing modules: preprocessing, CST segmentation, classification, parameters tuning and CST scoring (Fig. 2).

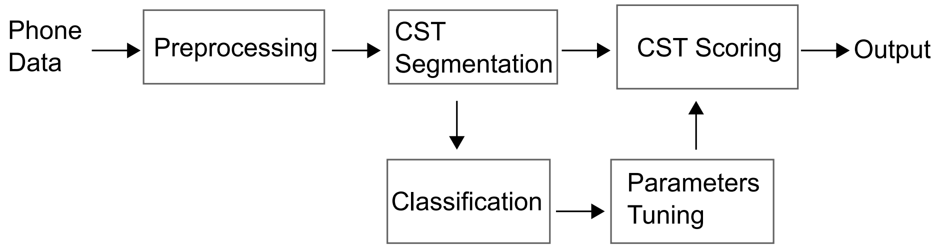

Fig. 2: Architecture of the proposed software system which is comprised of five data processing modules. For this study, phone data was analysed offline using custom-built Python scripts on a standard laptop. The long term goal is to run the architecture on mobile platforms for real-time CST analysis.

Module 1: Preprocessing
The first and last five seconds of the recording were cut out to remove experimenter activity while placing and taking the phone in and out of the phone holder. All recordings were then low pass filtered using 2^{nd} order, zero phase Butterworth filter with 10 Hz cut off frequency.

Module 2: CST segmentation
To detect beginning and end of the CST activity, a simple change detection algorithm was developed. Assuming that participants were sitting still (no acceleration or deceleration) before and after the CST activity (at least for few seconds), the algorithm looked for high amplitude acceleration and deceleration points to detect the onset and duration of the CST activity. Briefly, the algorithm divided the data into small segments using a moving window (half second with 50% overlap) and calculated maximum and minimum amplitudes of each segment. If these amplitudes were above an amplitude threshold, the segment was marked as "CST is on", otherwise it was marked as "CST is off". The onset and duration of the CST activity was estimated using the first and last "CST is on" segments. If the estimated duration was less than 28 seconds (this happened in few trials when participants could not finish the test), the recording was flagged for further inspection. If it was confirmed by the human expert that the CST activity was aborted early, the recording was marked as "incomplete" and excluded from the analysis.

Module 3: Classification
Two tree-based ensemble classifiers (XGBoost [12]) were created in Python using scikit-learn machine learning library: one for participant grouping (three classes: young, old and PD) and one for CRS recognition (two classes: flexion and MT).

Feature extraction The CST activity data was divided into three seconds segments with 50% overlap. The three second window size was chosen based on the minimum CST cycle observed in the entire data set (a similar value was previously reported for older adults [11, 13]). From each segment, time domain (max, min, std, mean, skew and kurtosis) and frequency domain (dominant frequency, amplitude and number of zero crossings) features were extracted. In total, the classification data set had 743 segments (observation points), and each segment had 54 output variables: 9 features $\times$ (3 accelerometer channels + 3 gyroscope channels).

Training and testing The original data set was not balanced. We had more segments from old participants (553) than young (97) and PD participants (93). We also had more CST cycles with MT strategy (162) than flexion strategy (581). Oversampling method from "imblearn" Python library was employed to increase the number of samples in minority classes. After oversampling the dataset included 1659 segments (from three participant groups) and 1162 CST cycles (from two CRS). This data set was split into training (80%) and testing (20%) sets using a stratified strategy (i.e., number of classes were always equal in testing and training sets). The classifiers were trained using the following parameters: maximum depth = 10, number of estimators = 100, learning rate = 0.3. Training was repeated 10 times by randomly shuffling the data. To evaluate how much accelerometer and gyroscope sensors contributed to the overall performance, two additional classifiers were trained and tested for each classification problem: one classifier using only accelerometer features and one classifier using only gyroscope features.

The performance of the classifiers were evaluated on test sets using precision, recall and f1-score. The confusion matrix (one per classifier) was also visualised to investigate incorrectly classified instances.

Module 4: Parameters tuning

Knowing who was taking the CST and which CRS they used was important to tune the parameters of the CST scoring module. To estimate CST cycle, the CST scoring module employed a peak detection algorithm originally designed for estimating heel strike and toe off time points during walking [14]. The performance of the peak detection algorithm depended on two key parameters (e.g., amplitude threshold and time threshold). Our preliminary work attempted to detect CST cycles using default values but this approach was found to have poor performance with high false positives (FP) and false negatives (FN) (e.g., FP = 20% and FN = 10% in old adults using MT strategy).

This was due to several reasons. First, the average CST cycle ($1-3$ seconds) was significantly longer than the average step cycle ($0.3-0.8$ seconds). Second, participants accelerated and decelerated more during CST than walking (at least in the vertical direction). Third, multiple prominent peaks were generated during each CST cycle.

We also noticed that acceleration profiles varied between the two CRS (for instance, the acceleration range, maximum acceleration $-$ minimum acceleration) was bigger in MT than in flexion Fig. 1. Similarly, the CST cycle varied significantly across participant groups: $0.98-1.3$ s (young), $1.5-2.6$ s (old) and

$2.4 - 3.5$ s (PD). This information was used while tuning the parameters of the peak detection algorithm. For instance, a PD participant using flexion strategy had lower amplitude and higher time threshold than young participant using MT.

Module 5: CST scoring
CST scoring module used the peak detection algorithm to detect CST cycles. The final CST score was predicted by counting the number of cycles. For each cycle, the range of vertical and forward accelerations, as well as the range of trunk angular velocity while leaning forward were also computed.

The predicted number of CST repetitions were compared to the actual number of repetitions obtained from the ground truth video recordings. The mean absolute difference between predicted and actual CST cycle period was also calculated.

2.3 Statistical tests

The average CST cycle completion times were compared between participant and CRS groups. To evaluate whether group means were statisticaly different from each other, non-parametric analysis of variance (ANOVA) using post-hoc conover test (with adjusted p-value) was employed.

3 Results

3.1 Participant and CRS Classification

When accelerometer and gyroscope features are used together, the XGBoost classifier had a classification accuracy of 97.8% (participant group) and 96.4% (CRS group). We also made an attempt to train a classifier using features from one sensor at a time. For participant classification, the performance of the classifier using gyroscope features was 90% for PD, 81% for young, 93% for old group. The performance of the classifier using accelerometer features was better: PD=100%, youn=91% and old=96%, but not as good as the classifier using combined features: PD=100%, young=99% and old=97% with precision =0.99, recall=0.98 and f1-score=0.98. For CRS classification, the performance of the classifier with combined features was 94% for flexion and 99% for MT strategies with precision, recall and f1-score of 0.97. The confusion matrix for each classifier (Fig. 3) suggests that there were very few classification errors between classes, the majority of these occurred between old and young subjects (Fig. 3).

3.2 CST characterisation based on CRS

In total, 591 CST cycles with flexion (208, pd=72, young=50, old=86) and MT (383, pd=41, young=108, old=234) strategies were investigated. Our previous

(a) Participant confusion matrix (b) CRS confusion matrix

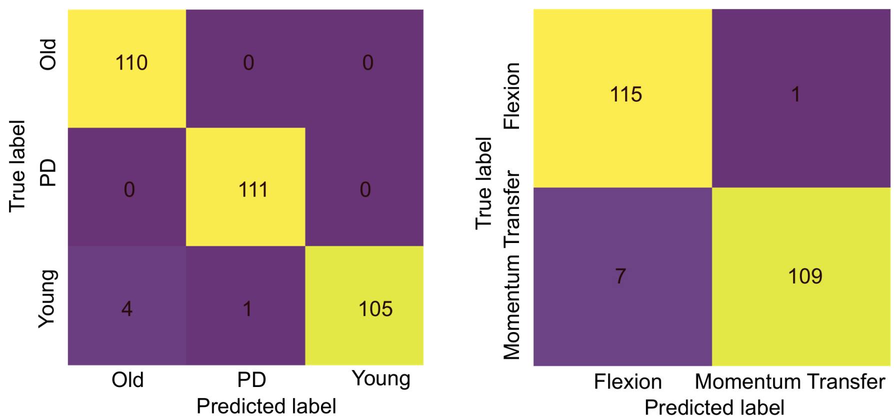

Fig. 3: Confusion matrices of XGBoost classifier on testing data: a) participant classification. b) CRS classification.

adaptive peak detection algorithm without participant and/or CRS information, only reached 71.2% CST cycle detection accuracy. With parameter tuning, where threshold values varied between 0.25 and 0.8 based on participant and CRS information, the accuracy increased to 96.4%. Similarly, the mean absolute difference in CST cycle time estimation between camera (ground truth) and phone was: PD (flexion = 50ms, MT=60ms), old (flexion=40ms, MT=20ms), and young (flexion=30ms, MT=25ms). Participant using flexion strategy had significantly higher CST cycle time than participants using MT strategy ($p < 0.01$) (Table 2).

PD subjects with flexion strategy had 6.4% higher cycle time than older adults and 36.3% than young subjects using the same strategy. PD subjects with MT strategy (these subjects were also ON medication) had 34.6% lower cycle time than PD subjects with flexion strategy. The time difference between MT and flexion was up to 40.7% in young and 44.8% in older adults. The system was able to detect the number of CST counts correctly in 98.7% cases with mean difference of count between camera and phone was less than one count.

Furthermore, participants with MT strategy had higher vertical acceleration than participants with flexion strategy. In contrast, participants with flexion strategy had higher turning velocity that the participants with MT strategy.

4 Discussion

Adaptation of peak detection algorithm to different CST rising strategies: CRS varied across participants and these variations caused false detection or failure to capture peaks by our initial peak detection algorithm previously proposed in [14]. As detection of peaks is critical for automating manual CST

Table 2: CST scoring using phone data. CST cycle time and number of repetitions are reported as mean ± standard deviation, whereas acceleration and angular velocity range are reported using min and max values.

Parameters	PD		Old		Young	
	Flexion	MT	Flexion	MT	Flexion	MT
No. of repetitions	10.3±2.4	15.5±2.5	9.5±2.4	20.3±3.9	16.6±1.7	27±4.6
CST cycle time (s)	2.81±0.6	1.85±0.3	2.63±0.8	1.45±0.4	1.79±0.2	1.06±0.2
Vertical acc. range (m s^{-2})	(-2.2, 10.6)	(-1.6, 15.9)	(3.8, 17.5)	(1.7, 22.3)	(1.8, 20.1)	(2.1, 22.8)
Forward acc. range (m s^{-2})	(-3, 7.6)	(-2.8, 7.6)	(-4, 7.8)	(-3.2, 12.1)	(-5.9, 8.1)	(-3.9, 8.4)
Trunk angular velocity range (rad s^{-1})	(-0.5, 0.6)	(-0.4, 0.3)	(-0.7, 0.5)	(-0.5, 0.4)	(-1.1, 0.7)	(-0.3, 0.4)

scoring, in this study we fine tuned the thresholds of the peak detection algorithm based on CRS and improved accuracy CST scoring significantly. Automation of subject and CRS identification enabled to select appropriate time and amplitude thresholds which enabled filtering of peaks with almost the same amplitude with smaller time difference between adjacent peaks. The major failure was experienced in flexion where the variation in the signal was higher and postural transition is not overlapped. However in momentum transfer it was different where body flexing and extension were happening at the same time resulting in generating peaks of actual standing and sitting with clear difference of amplitude. This is the reason participants had shorter CST cycle time and higher repetitions as compared to flexion.

Comparison of CST cycle score to existing studies: Two studies looked the estimation of CST cycle time using smartphone: one focusing on stroke subjects with mean cycle time=4.09±0.07s) [13] and the other focusing on older adults (2.81±0.5s) [11]. In our study, the cycle time of older adults using flexion strategy was 6.4% lower than those reported in [11], whereas the cycle time of older adults using MT strategy was 48% lower. Our cycle time results on flexion strategy aligns well with the existing literature possible because stroke and older participants are more likely to employ flexion strategy.

5 Conclusion

This study provides groundwork for employing smartphone-based data processing framework for automatic end-to-end analysis of CST in free living community settings. This framework operates in two tiers. First, it employs gradient boosting classifier driven by time and frequency domain features to recognise participant group (PD, old, young) and their CRS (flexion or MT). The classification accuracy of the system in both tasks was above 95%. Second, it selects the optimal time and amplitude threshold for the peak detection algorithm (based on participant group and CRS) to improve CST scoring. This adaptive approach had detected 99% of the CST repetitions and the cycle time estimation error was less than 60 ms in all data sets.

Future work The performance of proposed approach and its generalisation ability to unseen data needs further evaluation using larger data sets. We are also working towards developing an intelligent algorithm to detect and characterise anomalies during CST. Finally, we have an ongoing work investigating how CRS varies among stroke patients with low and high mRS (modified rankin scale) scores [15], and improving our smartphone system to characterise gait in outdoor environments [16].

Acknowledgements

This work is supported by Aberystwyth University (Presidential PhD scholarship to AS), Welsh Government (Sêr Cymru Cofund Fellowship, grant No. 663830-AU167, to OA, and Sêr Cymru Tackling Covid-19, grant No. 009 to FVP and OA, Health and Care Research Wales (PhD Studentship to FVP and OA), Public Health Wales - Stroke Implementation Group (Stroke Research, Innovation and Education fund to FVP and OA) and European Commission (H2020-MSCA-RISE-2019, grant No. 873178, to OA).

References

1. Lord, S.R., Murray, S.M., Chapman, K., Munro, B. and Tiedemann, A., 2002. Sit-to-stand performance depends on sensation, speed, balance, and psychological status in addition to strength in older people. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 57(8), pp.M539-M543.
2. Jones, C.J., Rikli, R.E. and Beam, W.C., 1999. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Research quarterly for exercise and sport, 70(2), pp.113-119.
3. Köpke, S. and Meyer, G., 2006. The Tinetti test. Zeitschrift für Gerontologie und Geriatrie, 39(4), pp.288-291.
4. Blum, L. and Korner-Bitensky, N., 2008. Usefulness of the Berg Balance Scale in stroke rehabilitation: a systematic review. Physical therapy, 88(5), pp.559-566.
5. Diao, Y., Lou, N., Liang, S., Zhang, Y., Ning, Y., Li, G. and Zhao, G., 2021. A novel environment-adaptive timed up and go test system for fall risk assessment with wearable inertial sensors. IEEE Sensors Journal, 21(16), pp.18287-18297.
6. Palmer, E., 2015. 10-meter walk test. EBSCO: Clinical Information Systems, pp.10-15.
7. Scarborough, D.M., McGibbon, C.A. and Krebs, D.E., 2007. Chair rise strategies in older adults with functional limitations. Journal of Rehabilitation Research & Development, 44(1).
8. Lummel, R.C., 2017. Assessing sit-to-stand for clinical use.
9. Mazzà, C., Benvenuti, F., Bimbi, C. and Stanhope, S.J., 2004. Association between subject functional status, seat height, and movement strategy in sit-to-stand performance. Journal of the American Geriatrics Society, 52(10), pp.1750-1754.
10. Alexander, N.B., Schultz, A.B. and Warwick, D.N., 1991. Rising from a chair: effects of age and functional ability on performance biomechanics. Journal of gerontology, 46(3), pp.M91-M98.

11. Cerrito, A., Bichsel, L., Radlinger, L. and Schmid, S., 2015. Reliability and validity of a smartphone-based application for the quantification of the sit-to-stand movement in healthy seniors. Gait & posture, 41(2), pp.409-413.
12. Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
13. Merchán-Baeza, J.A., González-Sánchez, M. and Cuesta-Vargas, A.I., 2018. Using smartphones to collect quantitative data on lower limb functionality in people who have suffered a stroke. Journal of Stroke and Cerebrovascular Diseases, 27(12), pp.3555-3562.
14. Sher, Arshad and Langford, David and Dogger, Einar and Lunn, Luke and Schröder, Maresa and Hamidinekoo, Azam and Arkesteijn, Marco and Shen, Qiang and Jansen, Thomas and Zwiggelaar, Reyer and Tench, Helen and Jones, Sarah and Jones, Phil and Povina, Federico, Villagra and Akanyeti, Otar, 2021. Automatic gait analysis during steady and unsteady walking using a smartphone (note: unpublished paper)
15. Meryem Şahin Erdoğan, Esra Sümer, Federico Villagra, Esin Öztürk Işık, Otar Akanyeti and Hale Saybaşılı, 2022, September.Predicting Modified Rankin Scale Scores of Ischemic Stroke Patients using Radiomics Features and Machine Learning. In UK Workshop on Computational Intelligence (to appear). Springer, Cham.
16. Bunker, M.T., Sher, A., Akpokodje, V., Villagra, F., Parthaláin, N.M. and Akanyeti, O., 2021, September. Towards Fuzzy Context-Aware Automatic Gait Assessments in Free-Living Environments. In UK Workshop on Computational Intelligence (pp. 463-474). Springer, Cham.