



Evaluation of the Skeleton Avatar Technique for Assessment of Mobility and Balance Among Older Adults

Sofia Backåberg^{1,2*}, Amanda Hellström¹, Cecilia Fagerström^{1,3}, Anders Halling⁴, Alisa Lincke⁵, Welf Löwe⁵ and Mirjam Ekstedt^{1,6}

¹ Department of Health and Caring Sciences, Faculty of Health and Life Sciences, Linnaeus University, Kalmar, Sweden,

² Sports Technology Research Lab, Faculty of Kinesiology, University of Calgary, Calgary, AB, Canada, ³ The Research

Section, Region Kalmar County, Kalmar, Sweden, ⁴ Department of Clinical Sciences, Center for Primary Health Care

Research, Lund University, Malmö, Sweden, ⁵ Department of Computer Science and Media Technology, Faculty of

Technology, Linnaeus University, Växjö, Sweden, ⁶ Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Background: Mobility and balance is essential for older adults' well-being and independence and the ability to maintain physically active. Early identification of functional impairment may enable early risk-of-fall assessments and preventive measures. There is a need to find new solutions to assess functional ability in easy, efficient, and accurate ways, which can be clinically used frequently and repetitively. Therefore, we need to understand how functional tests and expert assessments (EAs) correlate with new techniques.

Objective: To explore whether the skeleton avatar technique (SAT) can predict the results of functional tests (FTs) of mobility and balance: Timed Up and Go (TUG), the 30-s chair stand test (30sCST), the 4-stage balance test (4SBT), and EA scoring of movement quality.

Methods: Fifty-four older adults (+65 years) were recruited through pensioners' associations. The test procedure contained three standardized FTs: TUG, 30sCST, and 4SBT. The test performances were recorded using a three-dimensional SAT camera. EA scoring was performed based on the video recordings of the 30sCST. Functional ability scores were aggregated from balance and mobility scores. Probability theory-based statistical analyses were used on the data to aggregate sets of individual variables into scores, with correlation analysis used to assess the dependency between variables and between scores. Machine learning techniques were used to assess the appropriateness of easily observable variables/scores as predictors of the other variables included.

Results: The results indicate that SAT data of the fourth 4SBT stage could be used to predict the aggregated results of all stages of 4SBT (with 7.82% mean absolute error), the results of the 30sCST (11.0%), the TUG test (8.03%), and the EA of the sit-to-stand movement (8.79%). There is a moderate (significant) correlation between the 30sCST and the 4SBT (0.31, $p = 0.03$), but not between the EA and the 30sCST.

Conclusion: SAT can predict the results of the 4SBT, the 30sCST (moderate accuracy), and the TUG test and might add important qualitative information to the assessment of movement performance in active older adults. SAT might in the future provide the means for a simple, easy, and accessible assessment of functional ability among older adults.

Keywords: performance analysis, mobility, older adults, functional tests, balance

OPEN ACCESS

Edited by:

Maria Hägglund,
Uppsala University, Sweden

Reviewed by:

Mengxuan Ma,
University of Missouri, United States

Eling D. de Bruin,
Karolinska Institutet (KI), Sweden

Jordi Bernad,
University of Zaragoza, Spain

*Correspondence:

Sofia Backåberg
sofia.backaberg@lnu.se

Specialty section:

This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Computer Science

Received: 31 August 2020

Accepted: 23 November 2020

Published: 21 December 2020

Citation:

Backåberg S, Hellström A,
Fagerström C, Halling A, Lincke A,
Löwe W and Ekstedt M (2020)
Evaluation of the Skeleton Avatar
Technique for Assessment of Mobility
and Balance Among Older Adults.
Front. Comput. Sci. 2:601271.
doi: 10.3389/fcomp.2020.601271

INTRODUCTION

Maintaining mobility and physical activities of daily living among older adults has significant impact on quality of life, prolonging independent living, decreasing risk of falls, and reducing sedentary behavior (Rejeski et al., 2015; Aunger et al., 2018; Talarska et al., 2018). In the coming years, the proportion of people 65 years or older will increase dramatically, which implies a global challenge in several aspects (World Health Organization, 2015). This will require investments in promoting healthy aging and fundamental shifts in how we think about aging. The functional ability (i.e., mobility and balance) of older adults in everyday life is often limited by processes related to aging, such as gradual loss of muscle mass by 0.5–1% per year, as well as concomitant diseases. Levels of physical activity decrease with age (Lara et al., 2015; Stierlin et al., 2015), which has serious implications for the burden of chronic disease and mortality (Lee et al., 2012). Maintaining mobility despite disease can be crucial for being able to continue living at home, manage everyday life, and interact with the surrounding society. Strength, balance, and flexibility exercises are among the most effective strategies to counteract age-related decline of functional capacity and prevent falls among older adults (Paterson and Warburton, 2010; Dipietro et al., 2019). Physical activity is a modifiable behavior that contributes substantially to maintaining functional capacity and health (Lee et al., 2012). Thus, measures to prevent physical impairment and fall-related injuries for older adults are particularly important, with balance and mobility being essential aspects (Dipietro et al., 2019). Through regular assessment of functional ability, interventions could be initiated early, which might prevent mobility loss, improve quality of life, and prolong independent living among older adults. In community care, however, everyday rehabilitation and preventive work are not always prioritized alongside domestic care, and there is reason to believe that current functional tests (FTs) and similar assessment methods are insufficient.

Today, a range of tests and assessments of mobility and balance are available. Commonly, assessments of physical disability include performance-based (performance-oriented mobility assessment) and self-reported assessments (activities of daily living) (Cress et al., 1995). However, these are seldom routinely used, due to cost, the growing aging population, and the fact that an assessor, often a health professional, is needed to perform the tests. The consequences may be that physical activity and function of older adults become less visible and that early interventions are not being performed. For some diagnostic groups with chronic diseases, these tests can be exhausting or too challenging to carry out. The 30-s chair stand test (30sCST) is an example of this. The tests are commonly measured in time, counts, or distance. In the absence of expert assessments (EAs), crucial information about the qualitative aspects of the movement performance, such as compensatory movement

patterns, may be neglected, which may lead to inadequate health interventions and inadequate use of resources.

The advent of commodity three-dimensional (3D) sensor technology, e.g., the Kinect camera, has enabled efficient automated assessment of human movement. The Kinect camera technology is promising in detecting the risk of falls among older adults (Ejupi et al., 2015) assessing aspects of balance and postural control (Clark et al., 2015) and has been beneficial for the classification of the different stages of Parkinson disease related to freezing of gait (Dranca et al., 2018). Further research is needed to explore how the technique can be used to determine functional ability among older adults.

As the method still requires the use of the Kinect camera in a laboratory or having it installed at home, it excludes many people, especially those who have mobility difficulties (Ejupi et al., 2015). Also, our research was conducted using version 2 of the Kinect camera, which has been discontinued. However, the new version called Azure Kinect Development Kit has been released in 2019¹; it is easier to use. There are also alternative low-cost 3D camera systems, such as Orbbec's Astra Mini² and Intel's Real Sense³. These alternatives were tested against the Kinect version 2 and showed comparable tracking abilities (Hagelbäck et al., 2019b). Moreover, two-dimensional (2D) tracking software, such as PoseNet⁴, and the pose detection of Google's ML Kit⁵ exemplify another promising technology development. This type of software can be integrated in commodity mobile phones enabling skeleton avatar technique (SAT) to be both widespread and handy. Our own ongoing research maps the 2D SAT data of these software systems to 3D SAT data with high accuracy. In summary, the study presented here shows the predictive potential of SAT among older adults. On top of this feasibility study and based on the latest SAT development, there is a high potential of this technology facing a wider adoption in elderly care and even among elderly people.

Machine learning approaches map recorded movement instances to expert scores providing an automated assessment of the movements (Dressler et al., 2019a; Hagelbäck et al., 2019a). The so-called *skeleton avatar technique* (Dressler et al., 2019b) refers to the pipeline of hardware, software, and artificial intelligence components that records human movements with a 3D sensor, estimates the position of joints in each frame of the movement recording, and maps this information to a movement quality score.

Also, existing measures of mobility and balance among community-dwelling older adults have several limitations due to incompleteness, ceiling effects, and limited sensitivity to change and responsiveness (Lundin-Olsson, 2010; Pardasany et al., 2012, 2013). Furthermore, there is a lack of a common language among different professional groups regarding the assessment of balance and mobility, which can, for example,

¹<https://azure.microsoft.com/en-us/services/kinect-dk>

²<https://orbbec3d.com/astra-mini-series>

³<https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html>

⁴<https://github.com/tensorflow/tfjs-models/tree/master/posenet>

⁵<https://developers.google.com/ml-kit/vision/pose-detection>

Abbreviations: 30sCST, 30-s chair stand test; 4SBT, 4-stage balance test; CNN, convolutional neural network; EA, expert assessment; FT, functional test; MAE, mean absolute error; RNN, recurrent neural network; SA, self-assessment; SAT, skeleton avatar technology; TUG, Timed Up and Go.

hinder transitions between different care levels. Such difficulties when assessing mobility may reduce diagnostic sensitivity and the ability to capture improvements resulting from initiated interventions. Currently, a combination of several measures has to be used to encompass all aspects of functional ability (Berg and Norman, 1996; Dite and Temple, 2002). Thus, it is of significant importance to develop simple, inexpensive, and accurate assessment tools that are suited to the older adult's situation and can be used at a large scale in community care. As the first step in this development, the objective of this pilot study was to investigate the correlations between three FTs of mobility and balance [Timed Up and Go (TUG), 30sCST, 4-stage balance test (4SBT)], EAs, and the SAT.

MATERIALS AND METHODS

Study Design and Participants

This pilot study applied a cross-sectional design and was performed in purposely arranged and separate rooms at four different locations in the south of Sweden. Community-dwelling older adults (>65 years) were recruited via four pensioners' associations through emails/phone calls. In total, 54 older adults (38 females and 16 males) signed up for this study. All participants included in the study signed an informed consent form approved by the Swedish Ethical Review Authority (Dnr: 2019-02553).

Data Collection

Participants first completed a questionnaire with demographic information about their gender, age, weight, height, diagnosis, and symptoms, as well as a self-assessment (SA) of mobility and balance status (see Appendix 1 in the **Supplementary Material**). Twenty participants reported one or several medical diagnoses, with nine people reporting heart diseases, eight hypertension, four diabetes, and three reporting thyroid disease. Next, participants were instructed to perform three standardized FTs, the TUG, 30sCST, and 4SBT, which measure balance and mobility (Podsiadlo and Richardson, 1991; Rossiter-Fornoff et al., 1995; Jones et al., 1999). The FTs were performed in a controlled environment and supervised by a physiotherapist. Each test was recorded frontally with a Kinect sensor camera (Microsoft), with an infrared depth sensor of 512×424 pixels and an RGB camera resolution, $1,920 \times 1,080$ pixels. Its software development kit (SDK v2.0) computes 25 body joints (of which 13 were effectively used as discussed below), at a frequency of 30 Hz. The Kinect sensor camera was placed horizontally (no tilt angle) in 50-cm height. The participants were asked to stand in front of the camera before each test, so that the SAT could detect the person's full body (Dressler et al., 2019a; Hagelbäck et al., 2019a) in order to assess functional ability (mobility and balance). SAT data are 3D skeleton avatar sequences of the movements of the person performing FT.

Assessments

We assessed *functional ability* as an aggregate of *balance* and *mobility*, using three assessment approaches: an SA questionnaire, FTs, and an EA score of movement quality.

However, SA is subjective, and the relatively large margin of error when self-assessing physical activity levels is a well-known problem (Thyregod and Bodtger, 2016). SA should therefore be used together with other assessments. The FTs used in this study (TUG, 30sCST, and 4SBT) are standardized and objectively measured in time and require standardized settings and the involvement of a trained person during performance.

Expert Assessment

An experienced physiotherapist performed the EA of the sit-to-stand movement in this study, using a newly developed instrument for structured movement analysis of person transfer and mobility in physical activities of daily living (Backåberg et al., 2020). The instrument has been developed by an expert group of experienced physiotherapists, an occupational therapist, a researcher, and instructors within the field of safe person transfer and has been tested for face validity by a group of clinical physiotherapists. The instrument contains detailed descriptions of everyday life movements, focusing on the quality of the critical components of the movement performance. Performance is rated 0 = in accordance with the description, 1 = small deviation from the description, or 2 = large deviation from the description.

Mobility

Three mobility-related questions from the SA questionnaire and two standardized FTs (TUG and 30sCST) were used to measure the mobility of the participants. The mobility-related questions focused on sedentary behavior and levels of physical activity in daily life: time spent sitting or lying down during a day (scores 0–7, a higher score indicates a more sedentary lifestyle), time spent in physical activity during a week (scores 0–7, a higher score indicates more physical activity), and time spent exercising during a week (scores 0–7, a higher score indicates more exercise) (see Appendix 1 in the **Supplementary Material** and **Table 1**).

In the TUG test, the participants were asked to perform a sequence of movements: sitting in a chair with armrests, standing up, walking 3 m, turning around, going back, and sitting down again. The time needed to complete the test was measured. The TUG test has been shown to predict an elderly person's ability to walk independently (Podsiadlo and Richardson, 1991), and a score ≥ 14 s is associated with a higher risk of falls (Shumway-Cook et al., 2000). There is a categorization based on the time needed to perform the movement sequence, where ≤ 10 s is considered normal/no problems in mobility, 11–20 s = independence in movement, 21–29 s = large variation in functional ability, and more than 30 s = dependent/in need of assistance. The higher the score, the more difficulties. The test has shown good reliability and validity among healthy older adults (Podsiadlo and Richardson, 1991; Shumway-Cook et al., 2000). Previously reported psychometric properties of the TUG show high interrater reliability among community-dwelling older adults [Intraclass correlation coefficient (ICC) = 0.98] (Shumway-Cook et al., 2000), whereas another study shows that the test-retest reliability was moderate among older persons (Rockwood et al., 2000). The TUG has a high sensitivity (87%) and specificity (87%) and is able to identify elderly persons who are prone to fall. The discriminant analysis suggests that older adults

TABLE 1 | Demographic, self-assessment, and functional test data of participants ($n = 54$, missing 0–1.9%).

Variable	Men $n = 16$	Women $n = 38$	All $n = 54$
Age in years, mean (SD)	75.6 (3.2)	73.7 (4.7)	74.3 (4.4)
BMI, mean (SD)	26.2 (4.5)	25.9 (5.4)	26.0 (5.1)
Mobility			
Sitting and lying, hours/day, n (%)			
Never	0	0	0
1–3 h	4 (25.0)	3 (8.1)	7 (13.2)
4–6 h	9 (56.3)	22 (59.5)	31 (58.5)
7–9 h	1 (6.3)	8 (21.6)	9 (17.0)
10–12 h	1 (6.3)	4 (10.8)	5 (9.4)
13–15 h	1 (6.3)	0	1 (1.9)
Most all day	0	0	0
Physical activity, min/week, n (%)			
>300 min	8 (50)	20 (54.1)	28 (52.8)
150–299 min	5 (31.3)	7 (18.9)	12 (22.6)
90–149 min	1 (6.3)	5 (13.5)	6 (11.3)
60–89 min	0	1 (2.7)	1 (1.9)
30–59 min	0	3 (8.1)	3 (5.7)
<30 min	2 (12.5)	1 (2.7)	3 (5.7)
No time	0	0	0
Strenuous activity, min/week, n (%)			
>300 min	0	2 (5.4)	2 (3.8)
150–299 min	4 (25.0)	6 (16.2)	10 (18.9)
90–149 min	4 (25.0)	9 (24.3)	13 (24.5)
60–89 min	2 (12.5)	8 (21.6)	10 (18.9)
30–59 min	2 (12.5)	6 (16.2)	8 (15.1)
<30 min	3 (18.8)	3 (8.1)	6 (11.3)
No time	1 (6.3)	3 (8.1)	4 (7.5)
TUG test, n (%)			
No problems	14 (87.5)	36 (94.7)	50 (92.6)
Independence in movement	2 (12.5)	2 (5.3)	4 (7.4)
Large variation in functional ability	0	0	0
Dependent	0	0	0
30sCST, n (%)			
Below normal	3 (18.8)	6 (15.8)	9 (16.7)
Normal	9 (56.3)	15 (39.5)	24 (44.4)
Above normal	4 (25.0)	17 (44.7)	21 (38.9)
Balance			
Experienced difficulties with balance last 12 month, n (%)			
No difficulties	10 (62.5)	22 (59.5)	32 (60.4)
Difficulties	6 (37.5)	15 (40.5)	21 (39.6)
Number of falls last 12 months, n (%)			
No falls	12 (75.0)	30 (81.1)	42 (79.2)
One	1 (6.3)	6 (16.2)	7 (13.2)

(Continued)

TABLE 1 | Continued

Variable	Men $n = 16$	Women $n = 38$	All $n = 54$
Two	2 (12.5)	1 (2.7)	3 (5.7)
Three	0	0	0
More than three	1 (6.3)	0	1 (1.9)
4SBT, n (%)			
Stage 1 feet side by side 10 s	16 (100)	38 (100)	54 (100)
Stage 2 instep touches toe 10 s	16 (100)	37 (100)	53 (100)
Stage 3 tandem stand 10 s	12 (75.0)	28 (75.7)	40 (75.5)
Stage 4 one foot stand right 10 s	8 (50.0)	16 (43.2)	24 (44.4)
Stage 4 one foot stand left 10 s	8 (50.0)	18 (48.6)	26 (48.1)

who take longer than 14 s to complete the TUG have an increased risk of falls (Rockwood et al., 2000; Shumway-Cook et al., 2000).

In the 30sCST test, the participants were asked to rise from a chair repeatedly within 30 s. The test is used to assess the mobility and strength by timing the maximum number of stands from a chair in 30 s. The stand is performed with arms crossed over the chest and feet parallel. A score for the expected number of sit-to-stands, adjusted for age and gender, is provided. The number of stands is then categorized in three groups: *below normal*, *normal*, or *above normal*. The higher the score, the better the mobility. The test has good validity and reliability in measuring lower body strength in the elderly (Jones et al., 1999). Test–retest intraclass correlations of the 30sCST are high in both men (0.84) and women (0.92), indicating good stability of the measure. Moderate correlations between the test and leg-press performance suggest that 30sCST is a reasonably reliable and valid indicator of lower body strength in generally active, community-dwelling elderly. Construct validity is supported by the test’s ability to detect differences between various age and physical activity level groups (Jones et al., 1999).

Balance

Two balance-related questions in the SA questionnaire [if the person had experienced difficulties with their balance within the last 12 months (yes = 1, no = 0) and number of falls the last 12 months (score 0–5, the higher the score, the more falls)], together with one of the FTs (4SBT), were used to measure the balance of the participants.

The 4SBT is used to assess static balance (Rossiter-Fornoff et al., 1995). The participants are instructed to stand in four different positions that are progressively harder to maintain. First, the person is instructed to stand with their feet side-by-side. The next position is to place the instep of one foot, so it is touching the big toe of the other foot. The third position is a tandem stand, i.e., to put one foot in front of the other, heel touching toe. Lastly, the person tries to stand on one foot. How

long each position is held is measured in seconds; the hold should preferably exceed 10 s without the person moving his/her feet or needing support. The inability to maintain a tandem stand for 10 s has been associated with an increased risk of having a fall (Gardner et al., 2001). Test–retest of the 4SBT shows moderate correlation (0.66) in community-dwelling elderly. The test is also correlated to other measurements for balance in the same population (Rossiter-Fornoff et al., 1995).

Data Preprocessing

In the data preprocessing step, we applied the following filters and transformations to the FT results: (a) removing columns that had zero variance (all participants performed the exercise equally); (b) normalizing the test results using the (complementary) cumulative (sample) distribution function; (c) aggregating the four individual balance test results (two individual mobility test results, respectively) to a common balance (mobility) score using the joint cumulative (sample) distribution of the individual balance (mobility) scores. For each subject, the data transformations (b) and (c) computed a score expressing the probability of a performance worse than the subject's performance. The same transformation was applied to the SA and EA results.

The normalization (b) and aggregation (c) steps deserve some explanations and motivations: The cumulative distribution function (CDF) of a variable X is the probability that X will take a value less than or equal to x , i.e., $\text{CDF}_X(x) = P(X \leq x)$. It requires that X is measured at least on an ordinal scale, i.e., “less than” (\leq) is defined on X . For instance, balance test times induce an order; gender does not.

The distribution of X is, in general, unknown and can only be approximated numerically by observing a (representative, sufficiently large) sample X of the population X . Then, the empirical (or sample) cumulative distribution function (ECDF) is a good approximation of CDF. $\text{ECDF}_X(x)$ can be calculated as the relative frequency of observations in the sample X that are less than or equal to x , i.e., $\text{ECDF}_X(x) = |\{x|x \in X, x < x\}|/|X|$ with $|\cdot|$ the size of a set.

The number of participants in the described study is relatively small and may be biased. Hence, we cannot claim that we assessed a representative sample X of the population X in any of the FTs. Consequently, ECDF_X may not yet be a good approximation of CDF_X . However, our study shows the predictability of the (empirical) scores from SAT data. It is plausible that this predictability continues to hold for scores based on larger samples, as well. Then, our deep learning model would predict the normalized score based on SAT data, and ECDF^{-1} of this score, the actual variable values.

What is considered a sample set that is sufficiently large depends on the number of distinguishable variable values; we would like to observe each possible value at least once in the sample. This number is finite for discrete value domains, e.g., the number of squats that are possible within 30 s for 30sCST (0... <100), but also for physically continuous value domains due to discretization of the measurement method and the digital representation, e.g., the balance time in 10th of a second for each 4SBT stage (0... 100). However, in the latter case, this

number might become large, in general, e.g., when we would assess the balance time in microseconds (0... 10,000). Then, sufficiently large would become prohibitively large and expensive for practical studies. Consequently, for any reasonably practical sample size, ECDF would not be an injective but a *step* function, and its inverse ECDF^{-1} would provide the predicted *interval* of the variable values. This could either be accepted or avoided by using smoothing PDF/CDF estimations⁶.

In the present study, we did not apply any mitigation and accept the too small sample as a limitation: although our study shows the predictability of the empirical scores from SAT data, it is not the capable of predicting the actual value of a test.

(E)CDF nicely generalizes to multivariate distributions allowing to integrate different variables into one score. For instance, the results from all stages of the 4SBT (five variables, say $X_1, X_2, X_3, X_{4r}, X_{4l}$) can be integrated into a balance score using the *joint* CDF. Again, each variable needs to be measured at least on an ordinal scale. For our purpose of scoring, it also needs to be known whether large or small values are desirable. For instance, for each variable $X_1, X_2, X_3, X_{4r}, X_{4l}$, larger balance times (up to 10 s) are desirable. If small values are desirable, we use the complementary CDF defined as $\text{CCDF}_X(x) = P(X \geq x)$. As a consequence, regardless of whether large or small variable values are desirable, larger scores are always better than smaller, and 1 (resp. 0) is the best (resp. worst) possible score. The empirical complementary CDF is computed analogously to ECDF, i.e., $\text{ECCDF}_X(x) = |\{x|x \in X, x \geq x\}|/|X|$. Without loss of generality, we continue our discussion based on the (E)CDF and do not explicitly mention the variables requiring (E)CCDF scoring.

Unfortunately, the inverse of a joint CDF is not unique. In general, CDF^{-1} maps a score s to a *set* of vectors $\{v_1 \dots v_n\}$, each vector with positions for each variable. More precisely, $\{v_1 \dots v_n\} = \text{CDF}^{-1}(s)$ iff $s = \text{CDF}(v_1), \dots, s = \text{CDF}(v_n)$. The vectors $\{v_1 \dots v_n\}$ are not comparable; i.e., for any two of them, say v and v' , it is neither $v \leq v'$ nor $v \leq v'$. However, the set of vectors can be abstracted to value intervals of lower and upper values for each of the variables. For instance, for a concrete aggregated balance score, it can be stated that each of $X_1, X_2, X_3, X_{4r}, X_{4l}$ is within a concrete value interval.

Yet again, for representative, sufficiently large samples of the population, the joint ECDF approximates the joint CDF. Our deep learning model would predict the normalized joint ECDF score based on SAT data, and ECDF^{-1} of this score, the corresponding variable value intervals.

In summary, the normalized scores $s(v)$ of any measured/observed value v computed step (b) can be interpreted as the (sample) probability of finding a worse or equal value in the (sample) population. What “worse” means depends on the interpretation of the respective test. For example, lower values of physical activity (in minutes/week) are worse, whereas higher values of sitting and lying down (hours/day) are worse. To compensate for the different interpretations of values, we used the cumulative (sample) distribution function if high values were encouraged for a test, whereas we used the complementary

⁶For instance, the Parzen–Rosenblatt window method, https://en.wikipedia.org/wiki/Kernel_density_estimation.

cumulative (sample) distribution function if low values were encouraged. As a result, all scores were normalized to between 0 and 1, and high scores were always better than low scores. For details of the normalization method, we refer to Ulan et al. (2019).

The SAT recorded movement sequences of the subjects' 3D joint positions. The Kinect camera used identified 25 such joints. Because of the low reliability of the other joints, we used only the following 13: head, left/right shoulder, left/right elbow, left/right wrist, left/right hip, left/right knee, and left/right ankle.

Each recorded movement is a sequence of frames. A frame is a record of features. It describes the body posture at a specific point in time during the recorded movement. A feature is called direct if it is directly measured by the 3D camera and indirect if it is computed from direct features or other indirect features. The direct features include the x , y , and z coordinates of 13 skeleton joints. Indirect features include the angles between different limbs and angles between limbs and the axes of the 3D coordinate system.

We conducted each machine learning experiment twice, once with the standardized features and once with the raw (direct and indirect) features.

We tested both the uncut sequences, including some frames where subjects were getting into position before starting the movement, and the sequences cut to encompass only the actual movement from start to finish. For the TUG tests, we always cut at the turning point.

In accordance with a standard technique in machine learning, we used data augmentation to artificially increase the number of training and test sequences. In general, data augmentation increases the amount of training data, e.g., by adding slightly modified copies of existing data, which reduces overfitting when training a model (Shorten and Khoshgoftaar, 2019). Specifically, we stretched each frame in the x and y directions by the same constant factors around 1, and we rotated each frame around the y axis by the same constant angle around 0 degrees. Cascading these transformations led to an increase of the number of sequences for machine learning by a factor of about 1,000.

Statistical Analysis

SPSS 26.0 (IBM Corp., Armonk, NY, USA) was used for descriptive statistics. Pearson correlation analysis was conducted using MATLAB version R2020a (Massachusetts, USA)⁷. Significance was set at $p < 0.05$. Pearson correlation coefficient r was used to determine the dependencies between three assessment approaches for mobility and balance. The correlation results were interpreted as low ($r < 0.30$), moderate ($0.30 \leq r < 0.60$), or high ($r \geq 0.60$).

Machine Learning/Deep Learning

In general, machine learning uses predictors (features) $X = [X_1, \dots, X_p]$ to predict or infer a response (output) Y . It assumes a functional relationship, $Y = f(X_1, \dots, X_p) + e$, with e being an irreducible random error and f representing the systematic

information that X provides about Y . Statistical/machine learning tries to estimate functions f minimizing the *reducible* error. More precisely, to predict $Y = f(X_1, \dots, X_p) + e$, machine learning calculates an estimator function $Y' = F(X_1, \dots, X_p)$ and uses Y' as a predictor of Y .

If the estimator function F is accurate; i.e., the error between the actual response Y and its predictor is always small, machine learning can answer questions such as what is the expected value y of Y given the values $x = [x_1, \dots, x_p]$ for the predictors $X = [X_1, \dots, X_p]$. Moreover, if the estimator function F is sufficiently simple, machine learning can also give answers to questions such as the following: Which predictors are associated with the response? What is the relationship between the response and each predictor? Can the relationship between the response and each predictor be adequately summarized using a known type of function, e.g., linear? Unfortunately, there is a trade-off between prediction accuracy and estimator interpretability. For further details, we refer to machine learning textbooks such as James et al. (2013).

There are different ways of formalizing the reducible error. We selected the mean absolute error (MAE) for both learning F and assessing its accuracy on the training and test data, respectively. MAE is defined as the arithmetic mean of the absolute difference $|y - y'|$ for each actual response $y \in Y$ and its corresponding predictor value $y' \in Y'$ in the training and test data, respectively.

In our experiments, we use *deep learning* approaches that are known to trade off interpretability against accuracy. Here, deep learning approximates a function mapping sequences of 3D joint positions (preprocessed as described earlier) to the different SA, FT, and EA scores. The input shape depends on the number of indirect features and the number of frames in the shortest sequence; both varied in the different setups. As the responses, i.e., the different scores, are normalized to 0–1, the MAE is also between 0 (no error) and 1 (theoretical maximum). We interpreted the machine learning results, i.e., the accuracy of the trained predictor, as good ($MAE < 10\%$), moderate ($10\% \leq MAE < 20\%$), or bad ($MAE \geq 20\%$).

Our experiments applied standard neural network technology (Goodfellow et al., 2016) implemented in Python 3 using the Tensorflow framework (Abadi et al., 2015).

Architecture

We tested three principally different neural network architectures with roughly the same number of parameters to learn.

1. A dense network with three dense layers of 128, 64, and 32 neurons, respectively, all activated with a rectified linear unit (ReLU), and an output layer with a single output (the score) activated with a sigmoid activation function.
2. A convolutional neural network (CNN) with three one-dimensional (1D) convolutional layers with a depth of 128, 64, and 32 neurons, respectively, and all followed by a 1D maximum pooling layer of size two and activated with an ReLU, followed by an output layer, as in 1.
3. A recurrent neural network (RNN) with three long short-term memory layers of 32 neurons each, followed by an output layer, as in 1.

⁷The Mathworks Inc., Natick, Massachusetts: MATLAB version 9.8.0.1323502 (R2020a).

In all architectures, we used either dropout, with a rate of 0.5 in the first layer, or kernel and activation regularization (L2 norm, penalty of 0.001) of the first two layers.

Training

We randomly split the original sequences into about 90% training data and 10% test data. We did not mix the augmented sequences. All transformed training (test) data sequences remained in the training (test) data set. We did not separate test and validation data.

For training, we used the MAE as the loss function. We trained the networks with a minibatch size of 128 data points for 500 epochs. We used early stopping if the validation loss (MAE) did not decrease for the latest 50 epochs. The whole machine learning process is summarized in **Figure 1**.

We used the Tensorflow default weight initialization (Glorot uniform initializer) for all layers. It draws samples from a uniform distribution within $[-limit, limit]$, $limit = \sqrt{6 / (in\ degree + out\ degree)}$, and in (out) $degree$ the number of predecessors (successors) of a neuron.

We used “Adam” as the gradient-based weight optimization strategy for the dense and the convolutional networks, and “RMSprop” for the RNNs. Both approaches are implemented in Tensorflow, and we applied the provided default hyperparameters (learning rate, etc.).

In general, we avoided fine-tuning of hyperparameters. The explicitly set training parameters were initially chosen by experience and then only minimally adapted after a visual inspection of the learning history in some few initial tests (as reported). For most hyperparameters, we chose the default settings of the Tensorflow framework. The rationale behind this approach was that the goal of the present study was to principally show the predictive power of SAT data for FT scores. The small sample size alone prohibited aiming for optimal prediction models or minimal training times. This will become future work when large and representative samples are available.

Machine learning approximates a predictor function mapping predictor values (here, SAT frame sequences) to response values (here, the different normalized values, e.g., the time in seconds that a subject was able to stand on one leg). The results reported are the MAEs of the predictor functions applied on the test data using cross-validation. In detail, we

1. added features to the SAT frame sequences, such as angles between adjacent limbs, or skipped this step;
2. cut the SAT frame sequences to encompass the period between start and stop of the actual exercise, or skipped cutting;
3. augmented the resulting SAT frame sequences;
4. standardized the SAT frame sequence data, or skipped standardization;
5. normalize the observed values, cf. preprocessing (b);
6. aggregated these values to a score, cf. preprocessing (c);
7. copied the scores such that each SAT frame sequence transformed in Rejeski et al. (2015) got the same score as its original;
8. performed 10-fold cross-validation that iterated 10 times through the deep learning (step 9); and
9. performed deep learning on each fold.

The cross-validation step 8 randomly split the preprocessed predictor and response data into 10-fold. Cross-validation was iterated 10 times through the deep learning step 9, each time choosing a new fold as the test data and using the remaining folds as the training data. The overall result is the average over the 10 computed MAEs from each iteration. In each iteration, step 9 learns an estimator function F for the training data, mapping the contained predictor to response data. Its MAE is then computed on the test data.

Because of the high computational effort, cross-validation was only performed on promising combinations of preprocessed data and neural network model. Predicting FT and EA scores from balance SAT data using the CNN and the RNN based models gave promising results, i.e., good accuracy for predicting two of three FTs (4SBT, 30sCST) and the EA and moderate accuracy for the

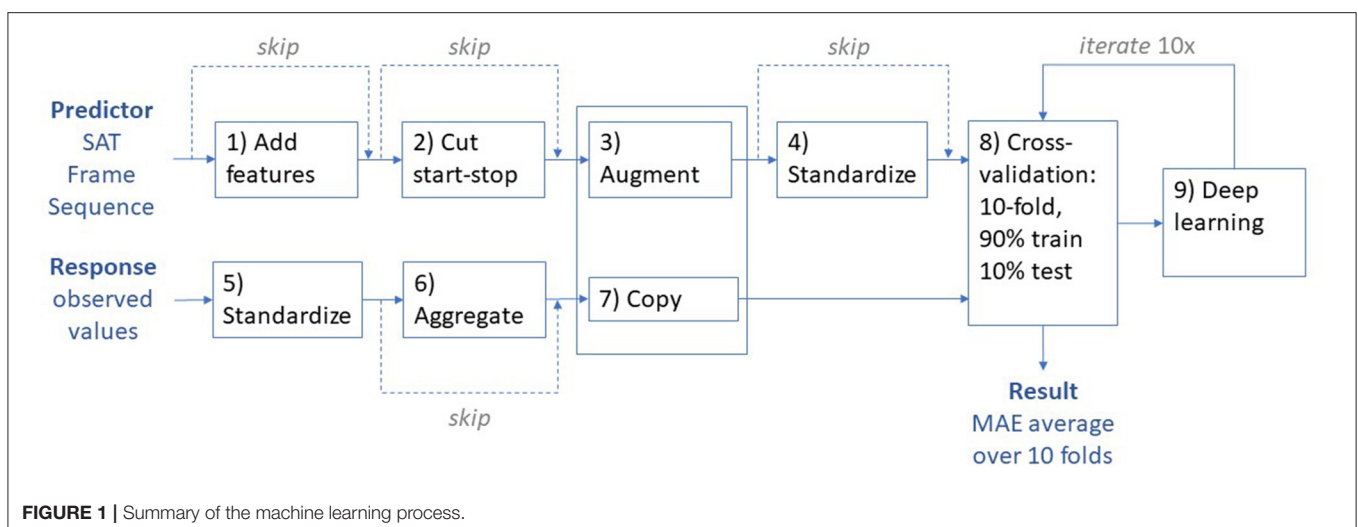


FIGURE 1 | Summary of the machine learning process.

third FT (TUG), as we will document in the following section. Therefore, we cross-validated these predictor models.

RESULTS

Summary of Collected Data

Table 1 presents a summary of the collected data for the FTs (TUG, 30sCST, 4SBT) and the SA questionnaire. Most participants performed well or very well on the tests. Twenty-eight persons (52.8%) carried out leisure-time physical activities for more than 5 h/week. All but four persons (7.5%) stated that they performed moderate to high levels of physical activities every week. Six persons (11.3%) reported that they spent 10–15 h sitting or lying down every day. Approximately half of the sample (52.8%) had an exercise program that they followed. How many times a week they performed the program varied, as did the length of the programs. Of those who followed an exercise program, 65.5% had a program that was 30 min or longer. Participants were also asked if they considered themselves physically active; 71.7% did. Eleven persons (20.8%) reported having had a fall one or several times during the last 12 months. Twenty-one persons (39.6%) stated that they had experienced difficulties with their balance within the last year (**Table 1**).

All study participants managed the TUG test within 20 s, indicating independent walking. According to the 30sCST, 45 persons (84.9%) in the present study had muscle strength as anticipated or better given their gender and age. All study participants managed the first stage of the 4SBT, i.e., standing for 10 s with feet side-by-side. The third stage of the 4SBT, i.e., the “tandem stand,” standing with one foot in front of the other, showed that 13 persons (24.5%) in the sample might be at risk of having a fall.

Correlation of Mobility and Balance Assessments Between FTs, SA, and EA

There was a moderate (significant) correlation between the 30sCT and the TUG test, and between the third stage of the 4SBT (4SBT3), i.e., the tandem stand position, and the fourth and final stage of the 4SBT (4SBT4), i.e., the one-foot standing position, on the right or left foot, respectively. No significant correlation was found between the second stage of the 4SBT (4SBT2), i.e., one foot placed with the toes at the insole of the other foot, and the other stages of the 4SBT, the TUG test, or the 30sCST. There was a high (significant) correlation between right and left foot for the 4SBT4 (**Table 2**). There was a moderate (significant) correlation between the functional balance test scores (FT balance), i.e., the 4SBT, and the SA scores that related to mobility (SA mobility) (**Table 3**). No significant correlation was seen between the 30sCST and EA.

Prediction of FT and EA Results Using the SAT

The SAT data-based neural network models were well able to predict the aggregated functional balance test score, the 30sCST, and the EA of the sit-to-stand movement. However, the models

TABLE 2 | Correlation coefficients of mobility and balance characteristics in FTs.

<i>r</i> (<i>p</i>)	TUG	30sCST	4SBT2	4SBT3	4SBT4, right
30sCST	0.55 (0.001)				
4SBT2	0.09 (0.536)	−0.04 (0.773)			
4SBT3	−0.03 (0.847)	0.18 (0.207)	−0.08 (0.576)		
4SBT4 right	0.22 (0.122)	0.25 (0.081)	0.12 (0.407)	0.34 (0.018)	
4SBT4 left	0.36 (0.011)	0.31 (0.029)	0.14 (0.318)	0.40 (0.004)	0.67 (0.000)

Bold font mean p < 0.05; TUG, Timed Up and Go; 30sCST, 30-s chair stand test; 4SBT2, second stage of the 4SBT, i.e., 10-s one foot placed with the toes at the insole of the other foot; 4SBT3, third stage of the 4SBT, i.e., 10-s tandem stand position; 4SBT4, fourth stage of the 4SBT, i.e., 10-s one-foot stand.

TABLE 3 | Correlation coefficients of aggregated mobility and balance scores in SA and FTs.

<i>r</i> (<i>p</i>)	SA balance	SA mobility	FT balance
SA mobility	0.2 (0.167)		
FT balance	0.18 (0.217)	0.40 (0.005)	
FT mobility	0.22 (0.134)	−0.05 (0.755)	0.19 (0.198)

Bold font mean p < 0.05. SA, self-assessment; FT, functional test.

could predict the result of the TUG test only with moderate accuracy (**Table 4**).

Moreover, based on the SAT data of the fourth and final stage of the 4SBT (4SBT4), i.e., 10 s, one-foot stand, the neural network models could well-predict the performance in the 30sCST and the EA of the sit-to-stand movement. They were able to predict the TUG test result only with moderate accuracy (**Table 5**).

To secure the results, we conducted 10-fold cross-validation on the predictions based on the SAT data of the 4SBT. We restricted the cross-validation to the CNN and RNN model variants. Cross-validation confirmed that SAT 4SBT4-based RNN models could predict the performance in the functional balance tests and the EA of the sit-to-stand movement. The RNN models consistently outperformed the CNN models. The RNN models were even able to predict the TUG test result with high accuracy. However, they could predict the 30sCST test result only with moderate accuracy (**Table 6**).

DISCUSSION

The results of this study indicate that the SAT-based data of the 10-s one-foot stand balance test (4SBT4) could be used to predict the results of all functional balance tests (MAE 7.82% cross-validated), the TUG test (MAE 8.03% cross-validated), and

TABLE 4 | MAEs for predicting FT results and EA scores using the corresponding SAT data.

	SAT 4SBT4 → FT balance	SAT 30sCST → FT 30sCST	SAT TUG → FT TUG	SAT 30sCST → EA sit-to-stand
MAE in %	2.31%	7.40%	16.22%	5.89%
Model/data	RNN/uncut, not norm., direct features	CNN/cut, normalized, direct features	CNN/uncut, normalized, all features	RNN/cut, normalized, all features

SAT, skeleton avatar technique; FT, functional test; 4SBT4, fourth stage of the 4SBT, i.e., 10-s one-foot stand; 30sCST, 30-s chair stand test; TUG, Timed Up and Go; EA, expert assessment; MAE, mean absolute error; RNN, recurrent neural network; CNN, convolutional neural network.

TABLE 5 | MAEs for predicting FT and EA scores using SAT balance data.

	SAT 4SBT4 → FT 30sCST	SAT 4SBT4 → FT TUG	SAT 4SBT4 → EA sit-to-stand
MAE in %	7.02%	18.21%	6.86%
Model/data	RNN/uncut, normalized, all features	CNN/uncut, normalized, all features	CNN/uncut, not norm., all features

SAT, skeleton avatar technique; FT, functional test; 4SBT4, fourth stage of the 4SBT, i.e., 10-s one-foot stand; 30sCST, 30-s chair stand test; TUG, Timed Up and Go; EA, expert assessment; MAE, mean absolute error; RNN, recurrent neural network; CNN, convolutional neural network.

the EA results of the sit-to-stand movement (MAE 8.79% cross-validated). They might be used to predict the results of the 30sCST (11.0% cross-validated). This first attempt to validate the SAT in relation to commonly used FTs in healthy and physically active older adults provides support to proceed with a larger sample of people with a varying degree of functional ability before the SAT can be used as an alternative method for assessing mobility and balance.

This study is the first step to outline the possibilities of using the SAT to obtain detailed, objective, and reliable information from simple and accessible functional assessment of balance and mobility. Muscular and functional asymmetries have been shown to represent a risk factor for falls among older adults, and a current review study outlined that symmetry in gait was correlated with better functional performance among older adults. Interventions to improve symmetry in movement patterns are therefore important (Guadagnin et al., 2019). Assessment of the qualitative aspects of movement performance, i.e., symmetry, as mentioned above, how the movement is initiated, how force is used to accomplish the movement, and how the movement is coordinated, requires a trained expert, e.g., a physiotherapist. Such assessment could be expensive and time-consuming and is therefore often overlooked. The qualitative aspects in the assessment of movement performance can be crucial, especially among older adults, because of their increasing need to use their physical resources optimally. This may, for example, play an important role in the ability to perform physical activities in daily life and could provide valuable information about the risk of falls. These aspects seem to be missing in the studied standardized functional assessment tests (TUG, 30sCST, and 4SBT), which has been confirmed in previous studies (Inkster

and Eng, 2004; Manckoundia et al., 2006). Including these aspects in the assessment of functional ability and in interventions targeting older adults is therefore important, as it may increase physical confidence, competence, and motivation for a safe, independent, and physically active life and reduce the risk of falls. The results from this pilot study indicate that the SAT has the potential to facilitate and supplement the clinically used FTs and may be a future solution to add qualitative perspectives to the assessment of functional ability.

Thirteen joints of the body were selected in the SAT analysis. Although the selected joints cover a large amount of the body segments, it is important to acknowledge that important movement segments (such as detailed movements of the feet or the neck) might still be missing in the overall analysis, which may be essential for the understanding of the whole movement performance. However, the SAT has the ability to add detailed information about relationships between multiple body segments in complex movement patterns, which is not possible to detect with the human eye. This information might be valuable in the clinical setting, i.e., for physiotherapists in movement assessment and evaluation. However, further development of SAT is needed to include all body segments. More research is furthermore needed to outline how the SAT can predict other kinds of movements, person transfers, and FTs and if the SAT can be used based only on a 2D video, which would imply greater accessibility for people to use the technique without expensive equipment and expert assistance. As we move into a period of increased population aging, everyday rehabilitation and preventive work in community care will not always be prioritized alongside domestic care. Thus, older adults may benefit from easy assessments of functional ability that can be used with the help of a nursing assistant or family caretaker in their own home. Furthermore, the SAT may create a greater possibility to detect physical impairment at an early stage, which is crucial in fall prevention and for the preservation of physical independence in aging.

Adults who have developed a lifelong understanding of the role of physical activity in healthy aging, who know about body movement skills and methods of improvement, may be more likely to sustain engagement in physical activity as an integral and meaningful part of their lifestyle in older age (Jones et al., 2018). Highly active persons of older age tend to use their resourcefulness to support their physical activity, which in turn contributes to their view of themselves as active. Barriers to being physically active in older age are all influenced by how older adults view themselves and how they are cognizant

TABLE 6 | Cross-validated MAEs for predicting FT and EA scores using SAT balance data; only recurrent neural networks (RNN).

	SAT 4SBT → FT balance	SAT 4SBT → FT 30sCST	SAT 4SBT → FT TUG	SAT 4SBT → EA sit-to-stand
MAE in %	7.82%	11.0%	8.03%	8.79%
Data	Uncut, not normalized, all features	Uncut, not normalized, all features	Uncut, not normalized, all features	Uncut, not normalized, all features

SAT, skeleton avatar technique; FT, functional test; 4SBT4, fourth stage of the 4SBT, i.e., 10-s one-foot stand; 30sCST, 30-s chair stand test; TUG, Timed Up and Go, EA, expert assessment; MAE, mean absolute error.

of and understand the social and physical environment and opportunities surrounding them (Jones et al., 2018). Although there is now strong evidence that regular physical activity is key to preserving physical function and mobility, which can delay the onset of major disability among older adults (Pahor et al., 2014; Dipietro et al., 2019), the majority of older adults do not achieve the recommended goals of physical activity (World Health Organization, 2015). Attitudes toward physical activity, lack of social support, feelings of being too old, and having few opportunities for physical activity in the surroundings are common barriers (Büla et al., 2011). There is vast opportunity for improvement in how to enhance physical literacy among older adults, which includes the motivation, confidence, and physical competence to achieve a physically active life (International Physical Literacy Association, 2015; Jones et al., 2018). In these efforts, the SAT may play an important role in facilitating assessments, providing feedback on movement performance, and improving physical competence, which in turn may contribute to the motivation and confidence to support and maintain physical activity throughout life.

In this study, EAs of the qualitative aspects of the movement performance were made only of the sit-to-stand movement. The duration of the sit-to-stand or stand-to-sit postural transition is commonly used for assessing function and strength of the lower extremities and can distinguish between older adults at low and high risk of falls. The sit-to-stand transition represents a complex motion that involves torques and forces on multiple joints (the trunk, hips, and knees), as well as requiring energy. However, it is argued that its duration is not sufficient to describe physical impairment in older adults (Inkster and Eng, 2004; Manckoundia et al., 2006). A model for optimality of the sit-to-stand movement has been developed that seems to be useful in detecting mobility changes (Madhushri et al., 2017). A fast sit-to-stand posture transition involves larger torques and greater wear on the body than a slow transition (Kerr et al., 1997). For each individual body constitution, there are movements/actions that provide optimal posture transition time. Physically fit persons are spontaneously very close to the optimal transition time, whereas older adults normally deteriorate in muscle strength and need to spend their resources wisely. Such persons might benefit from qualitative assessment of movement performance and functional ability, followed by tailored programs for exercise and mobility-oriented physical activity, to reach their optimal transition time. The results from the current study show that EAs of the sit-to-stand movement (Backåberg et al., 2020) did not correlate

well with the 30sCST, but could be predicted by the SAT. The reason for this is unknown, but it may indicate that there is a complex movement pattern that is not easily assessed and that there might be a non-linear association between EA and the sit-to-stand movement. The clinical assessments of functional ability may possibly be substituted by the SAT in future tools.

STRENGTH AND LIMITATIONS

A strength of the article is the effort to develop a novel approach of assessing physical functioning in older adults, such as early identification of functional ability using modern technology. Another strength for the development of SAT in the early stage as we are was the low internal dropouts and that the whole sample had the strength and ability to perform all the different tests included. However, some limitations are needed to pay attention to. This pilot study is based on a small sample ($n = 54$), which increases the risk of type II error. Another design limitation in this study is that no causal relationships between the variables can be identified with this set up. Although the participants were community-dwelling older adults (>65 years), they were recruited through a limited number of pensioners' associations, which might imply that this group included rather healthy, physically active, and dedicated older adults. Thus, the results of the SAT can primarily be generalized to this population of primarily healthy individuals and not fully represent the older population as a whole. The small and homogenous sample (with regard to mobility and physically active) was beneficial for the current developmental phase, but might have affected the results of the tests, as most participants performed very well on the tests, with small variations. This might have impacted on the SAT's ability to predict the results of the functional mobility tests (30sCST, TUG). The same applies to the functional balance tests (4SBT) and the EA of movement quality (EA).

To compensate for the relatively small sample size, we tested the significance in the correlation analysis. For the neural network learning experiments, we applied aggressive data augmentation and cross-validation, as detailed in *Data Preprocessing*. Still, more experiments with a larger sample are needed to confirm the results.

As use of the SAT in assessing older adults' mobility and balance is relatively new, more research is needed. In the next step, participants with more functional limitation variations should be included, to further evaluate the tests used. More sensitive tests would also be recommended.

The 3D technology was quite bulky and would need to become handier before any use in the caretaking practice can be recommended. Current activities aim at using 2D SAT-based mobile phone recordings. While the resulting skeleton avatar sequences contain even less information than the corresponding 3D-based sequences, they may contain enough information about the different FTs to yield relevant results.

The employed deep learning networks are hard to interpret for humans. The results merely show that the SAT sequences can provide systematic information about the FTs. To gain relevant insights into the dependency between the SAT sequences and the outcome of tests, other machine learning models should be used, features should be manually selected and deselected, and the prediction results of the different models should be compared.

CONCLUSION

Both in science and in clinical practice, there is a need to reduce the use of tests of functional ability that are difficult for patients to perform independently and replace with valid, simple, and accessible tools. In this study, we attempt to understand and verify what it is that we measure with FTs and how they correlate with new techniques. This study shows that the SAT may be a tool that is able to detect qualitative aspects in the assessment of movement performance, which seem to be missing in commonly used standardized functional assessment tests of mobility and balance (30sCST, TUG, and 4SBT). SAT was also shown to be able to a high extent predict the results of the 4SBT, TUG, and EAs of the sit-to-stand movement and, with some restrictions, the results of the 30sCST. However, this is the first attempt to use SAT as a functional assessment tool, and it needs to be investigated further, for example, how sensitive the SAT is to identify changes in physical activity levels and predict other aspects of functional ability. Research is also needed to investigate if SAT can identify changes in self-efficacy in movement performance, as well as securing the prediction of the 30sCST test in the older population.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Available online at: <https://arxiv.org/pdf/1603.04467.pdf> (accessed August 28, 2020).
- Aunger, J. A., Doody, P., and Greig, C. A. (2018). Interventions targeting sedentary behavior in non-working older adults: a systematic review. *Maturitas* 116, 89–99. doi: 10.1016/j.maturitas.2018.08.002
- Backåberg, S., Kimming, A., Kraft, P., and Reifeldt, K. (2020). *IRAF, an Instrument for Movement Analysis of Person Transfer and Mobility in Daily Living (in Swedish: Instrument för RörelseAnalys vid Förflyttning)*. Technical Report.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Swedish Ethical Review Authority (Dnr: 2019-02553). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AHe, CF, AHa, WL, and ME contributed to the conception and design of the research plan. AHe, SB, and WL were responsible for the acquisition of data. AHe, AL, and CF conducted the statistical analysis of the self-assessments, the functional tests, and the expert assessment. WL designed, implemented, and conducted the deep learning experiments. SB drafted the manuscript and completed the expert assessment. All authors contributed with writing different parts and provided a critical review of both intermediary and final drafts of the manuscript and approved the final draft prior to journal submission. All authors accept accountability for the parts of the work they have done.

FUNDING

Seed funding from the Linnaeus University Center for Data Intensive Sciences and Applications (DISA, Inu.se/disa) Developing the Skeleton Avatar camera Technique (SAT) as a rapid, valid and sensitive measurement of mobility in elderly persons.

ACKNOWLEDGMENTS

We are grateful to all the participants in the study, to the Linnaeus University Center for Data Intensive Sciences and Application and its High-Performance Computing Center for the financial support and the provision of the computing infrastructure, and to AIMO for developing a research version of their the SAT software used in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomp.2020.601271/full#supplementary-material>

- Berg, K., and Norman, K. E. (1996). Functional assessment of balance and gait. *Clin. Geriatr. Med.* 12, 705–723. doi: 10.1016/S0749-0690(18)30197-6
- Büla, C. J., Monod, S., Hoskovec, C., and Rochat, S. (2011). Interventions aiming at balance confidence improvement in older adults: an updated review. *Gerontology* 57, 276–286. doi: 10.1159/000322241
- Clark, R. A., Pua, Y. H., Oliveira, C. C., Bower, K. J., Thilarajah, S., McGaw, R., et al. (2015). Reliability and concurrent validity of the microsoft Xbox one kinect for assessment of standing balance and postural control. *Gait Posture* 42, 210–213. doi: 10.1016/j.gaitpost.2015.03.005
- Cress, M. E., Elaine, K. B., Schechtman, C. D., Mulrow, C. D., Fiatarone, M. A., Gerety, M. B., et al. (1995). Relationship between physical performance

- and self-perceived physical function. *J. Am. Geriatr. Soc.* 43, 93–101. doi: 10.1111/j.1532-5415.1995.tb06372.x
- Dipietro, L., Campbell, W. W., Buchner, D. M., Erickson, K. I., Powell, K. E., Bloodgood, B., et al. (2019). Physical activity, injurious falls, and physical function in aging: an umbrella review. *Med. Sci. Sports Exerc.* 51, 1303–1313. doi: 10.1249/MSS.0000000000001942
- Dite, W., and Temple, V. A. (2002). A clinical test of stepping and change of direction to identify multiple falling older adults. *Arch. Phys. Med. Rehabil.* 83, 1566–1571. doi: 10.1053/apmr.2002.35469
- Dranca, L., Ruiz de Mendarozketa, L. A., Goñi, A., Illarramendi, A., Gomez, I. N., Alvarado, M. D., et al. (2018). Using kinect to classify parkinson's disease stages related to severity of gait impairment. *BMC Bioinformatics* 19:471. doi: 10.1186/s12859-018-2488-4
- Dressler, D., Liapota, P., and Löwe, W. (2019a). "Data driven human movement assessment," in *Proceedings of the 11th KES International Conference on Intelligent Decision Technologies (IDT): Smart Innovation, Systems and Technologies Series Vol. 143*, eds I. Czarnowski, R. Howlett, and L. Jain (Singapore: Springer). doi: 10.1007/978-981-13-8303-8_29
- Dressler, D., Liapota, P., and Löwe, W. (2019b). "Towards an automated assessment of musculoskeletal insufficiencies," in *Intelligent Decision Technologies 2019 Smart Innovation, Systems and Technologies, Vol. 142*, eds I. Czarnowski, R. Howlett, and L. Jain (Singapore: Springer), 251–61. doi: 10.1007/978-981-13-8311-3_22
- Ejupi, A., Brodie, M., Gschwind, Y. J., Lord, S. R., Zagler, W. L., and Delbaere, K. (2015). Kinect-based five-times-sit-to-stand test for clinical and in-home assessment of fall risk in older people. *Gerontology* 62, 118–124. doi: 10.1159/000381804
- Gardner, M. M., Buchner, D. M., Robertson, M. C., and Campbell, A. J. (2001). Practical implementation of an exercise-based falls prevention programme. *Age Ageing* 30, 77–83. doi: 10.1093/ageing/30.1.77
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.
- Guadagnin, E. C., Barbieri, F. A., Simieli, L., and Carpes, F. P. (2019). Is muscular and functional performance related to gait symmetry in older adults? A systematic review. *Arch. Gerontol. Geriatr.* 84, 103899–103905. doi: 10.1016/j.archger.2019.103899
- Hagelbäck, J., Liapota, P., Lincke, A., and Löwe, W. (2019a). "The performance of some machine learning approaches in human movement assessment," in *Proceedings of the 13th Multi Conference on Computer Science and Information Systems (MCCSIS)*, eds M. Macedo, L. Rodrigues (Porto: IADIS Press), 35–42. doi: 10.33965/eh2019_201910L005
- Hagelbäck, J., Lincke, A., Löwe, W., and Rall, E. (2019b). "On the agreement of commodity 3D cameras," in *Proceedings of the 2019 International Conference on Image Processing, Computer Vision, and Pattern Recognition*, eds H. R. Arabnia, L. Deligiannidis, and F. G. Tinetti (Las Vegas, NV: CSREA Press), 36–42.
- Inkster, L. M., and Eng, J. J. (2004). Postural control during a sit-to-stand task in individuals with mild parkinson's disease. *Exp. Brain Res.* 154, 33–38. doi: 10.1007/s00221-003-1629-8
- International Physical Literacy Association (2015). *Canada's Physical Literacy Consensus Statement*. Vancouver BC: International Physical Literacy Association. Available online at: <http://physicalliteracy.ca/physical-literacy/consensus-statement/> (accessed July 29, 2020).
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer. doi: 10.1007/978-1-4614-7138-7
- Jones, C. J., Rikli, R. E., and Beam, W. C. (1999). A 30-s chair-stand as a measure of lower body strength in community-residing older adults. *Res. Q. Exerc. Sport* 70, 113–119. doi: 10.1080/02701367.1999.10608028
- Jones, G., Stathokostas, L., Young, B. W., Wister, A. V., Chau, S., Clark, P., et al. (2018). Development of a physical literacy model for older adults - a consensus process by the collaborative working group on physical literacy for older Canadians. *BMC Geriatr.* 18:13. doi: 10.1186/s12877-017-0687-x
- Kerr, K. M., White, J. A., Barr, D. A., and Mollan, R. B. (1997). Analysis of the sit-stand-sit movement cycle in normal subjects. *Clin. Biomech.* 12, 236–245. doi: 10.1016/S0268-0033(96)00077-0
- Lara, J., Cooper, R., Nissán, J., Ginty, A. T., Khaw, K.-T., Deary, I. J., et al. (2015). A proposed panel of biomarkers of healthy ageing. *BMC Med.* 13, 222–230. doi: 10.1186/s12916-015-0470-9
- Lee, I.-M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., et al. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 380, 219–229. doi: 10.1016/S0140-6736(12)61031-9
- Lundin-Olsson, L. (2010). Community-dwelling older adults with balance impairment show a moderate increase in fall risk, although further research is required to refine how balance measurement can be used in clinical practice. *Evid. Based Nurs.* 13, 96–97. doi: 10.1136/ebn.13.3.96
- Madhushri, P., Jovanov, E., Milenkovic, A., and Shtessel, Y. (2017). "A model based analysis of optimality of sit-to-stand transition," in *Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Seogwipo: IEEE). doi: 10.1109/EMBC.2017.8037339
- Manckoundia, P., Mourey, F., Pfitzenmeyer, P., and Papaxanthis, C. (2006). Comparison of motor strategies in sit-to-stand and back-to-sit motions between healthy and alzheimer's disease elderly subjects. *Neuroscience* 137, 385–92. doi: 10.1016/j.neuroscience.2005.08.079
- Pahor, M., Guralnik, J. M., Ambrosius, W. T., Blair, S., Bonds, D. E., Church, T. S., et al. (2014). Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. *JAMA* 311, 2387–2396. doi: 10.1001/jama.2014.5616
- Pardasaney, P. K., Latham, N. K., Jette, A. M., Wagenaar, R. C., Ni, P., Slavin, M. D., et al. (2012). Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. *Phys. Ther.* 92, 388–397. doi: 10.2522/ptj.20100398
- Pardasaney, P. K., Slavin, M. D., Wagenaar, R. C., Latham, N. K., Ni, P., and Jette, A. M. (2013). Conceptual limitations of balance measures for community-dwelling older adults. *Phys. Ther.* 93, 1351–1368. doi: 10.2522/ptj.20130028
- Paterson, D., and Warburton, D. (2010). Physical activity and functional limitations in older adults: a systematic review related to canada's physical activity guidelines. *J. Behav. Nutr. Phys. Act.* 7, 1–22. doi: 10.1186/1479-5868-7-38
- Podsiadlo, D., and Richardson, S. (1991). The timed "Up&Go": a test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* 39, 142–148. doi: 10.1111/j.1532-5415.1991.tb01616.x
- Rejeski, W. J., Rushing, J., Guralnik, J. M., Ip, E. H., King, A. C., and Maini, T. M. (2015). The MAT-sf: identifying risk for major mobility disability. *J. Gerontol. A Biol. Sci. Med. Sci.* 70, 641–646. doi: 10.1093/gerona/glv003
- Rockwood, K., Awalt, E., Carver, D., and MacKnight, C. (2000). Feasibility and measurement properties of the functional reach and the timed up and go tests in the Canadian study of health and aging. *J. Gerontol. A Biol. Sci. Med. Sci.* 55, M70–M73. doi: 10.1093/gerona/55.2.M70
- Rositter-Fornoff, J. E., Wolf, S. L., Wolfson, L. I., Buchner, D. M., and FICSIT Group. (1995). A cross-sectional validation study of the FICSIT common data base static balance measures. *J. Gerontol. A Biol. Sci. Med. Sci.* 50, M291–M297. doi: 10.1093/gerona/50A.6.M291
- Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6, 1–48. doi: 10.1186/s40537-019-0197-0
- Shumway-Cook, A., Brauer, S., and Woollacott, M. (2000). Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Phys. Ther.* 80, 896–903. doi: 10.1093/ptj/80.9.896
- Stierlin, A., De Lepeleere, S., Cardon, G., Dargent-Molina, P., Hoffmann, B., Murphy, M., et al. (2015). A systematic review of determinants of sedentary behaviour in youth: a DEDIPAC-study. *Int. J. Behav. Nutr. Phys. Act.* 12, 133–152. doi: 10.1186/s12966-015-0291-4
- Talarska, A., Tobis, S., Kotkowiak, M., Strugała, M., Stanisławska, J., and Wiecezorowska-Tobis, K. (2018). Determinants of quality of life and the need for support for the elderly with good physical and mental

- functioning. *Med. Sci. Monit.* 24, 1604–1613. doi: 10.12659/MSM.907032
- Thyregod, M., and Bodtger, U. (2016). Coherence between self-reported and objectively measured physical activity in patients with chronic obstructive lung disease: a systematic review. *Int. J. Chron. Obstruct. Pulmon. Dis.* 11, 2931–2938. doi: 10.2147/COPD.S16422
- Ulan, M., Ericsson, M., Löwe, W., and Wingkvist, A. (2019). “Multi-criteria ranking based on joint distributions - a tool to support decision making,” in *Proceedings of the 18th International Conference on Business Informatics Research*, eds M. Pankowska, K. Sandkuhl (Katowice: Springer), 74–88. doi: 10.1007/978-3-030-31143-8_6
- World Health Organization (2015). *World Report on Ageing and Health*. Luxembourg: World Health Organization.
- Conflict of Interest:** WL is a co-founder of AIMO (<https://aimo-fit.com/>), which developed the SAT software used in this study.
- The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Backåberg, Hellström, Fagerström, Halling, Lincke, Löwe and Ekstedt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.