Review

Subtask Segmentation Methods of the Timed Up and Go Test and L Test Using Inertial Measurement Units—A Scoping Review

Alexis L. McCreath Frangakis [1,*], Edward D. Lemaire [2] and Natalie Baddour [1]

1 Department of Mechanical Engineering, Faculty of Engineering, University of Ottawa, Ottawa, ON K1N 6N5, Canada
2 Faculty of Medicine, University of Ottawa, Ottawa, ON K1H 8M2, Canada
* Correspondence: amccr008@uottawa.ca

Abstract: The Timed Up and Go test (TUG) and L Test are functional mobility tests that allow healthcare providers to assess a person's balance and fall risk. Segmenting these mobility tests into their respective subtasks, using sensors, can provide further and more precise information on mobility status. To identify and compare current methods for subtask segmentation using inertial sensor data, a scoping review of the literature was conducted using PubMed, Scopus, and Google Scholar. Articles were identified that described subtask segmentation methods for the TUG and L Test using only inertial sensor data. The filtering method, ground truth estimation device, demographic, and algorithm type were compared. One article segmenting the L Test and 24 articles segmenting the TUG met the criteria. The articles were published between 2008 and 2022. Five studies used a mobile smart device's inertial measurement system, while 20 studies used a varying number of external inertial measurement units. Healthy adults, people with Parkinson's Disease, and the elderly were the most common demographics. A universally accepted method for segmenting the TUG test and the L Test has yet to be published. Angular velocity in the vertical and mediolateral directions were common signals for subtask differentiation. Increasing sample sizes and furthering the comparison of segmentation methods with the same test sets will allow us to expand the knowledge generated from these clinically accessible tests.

Keywords: timed up and go (TUG); L Test; wearable sensor; activity segmentation

1. Introduction

The Timed Up and Go test (TUG) and L Test are functional mobility tests that allow healthcare providers to assess a person's mobility during everyday life [1]. During the TUG, the person stands up from a chair, walks 3 m to a marker, turns around, walks back to the chair, and sits down. This test, therefore, consists of two 180° turns and 6 m of walking [1]. The L Test is a modified version of the TUG where the person stands up, walks 3 m to the first maker, turns 90°, walks 7 m towards a second marker, turns 180°, walks back towards the first marker, turns 90°, walks back to the chair, and sits down [2]. This incorporates 20 m of walking, with two 90° turns and two 180° turns. The L Test has been recommended to deal with a TUG ceiling effect for people with better mobility status [2] and further encompasses everyday movement since the two 90° turns force the person to turn both left and right, instead of just their preferred way [2].

Currently, to report a person's functional mobility, observations are made by a healthcare provider while the person completes the test, and the total test duration is measured using a stopwatch [3]. Subtasks are a way to further classify each movement in the test, in order to better understand the areas in which a patient may struggle [3]. The subtasks involved in the TUG and L Test include stand up, walk, turn, and sit down. In recent years, attempts have been made to automate data collection and identify subtask duration. This

will allow for a more concise profile of fall risk, dynamic balance, and agility [3]. Automated subtask segmentation often uses inertial measurement units (IMU), video recordings, or pressure-sensor data [3–5]. This review will focus on methods that use inertial data from devices such as IMU sensors or smartphones and perform subtask segmentation using only this inertial data.

IMUs collect data using accelerometers, magnetometers, and gyroscopes; however, not all parameters are used for segmenting each task. A physical representation of the parameters discussed in this report can be viewed in Figure 1. They include anteroposterior acceleration (AP_a), angular velocity (AP_ω), and rotation (AP_R); mediolateral acceleration (ML_a), angular velocity (ML_ω), and rotation (ML_R); and vertical acceleration (V_a), angular velocity (V_ω), and rotation (V_R). AP_R is sometimes referred to as roll, ML_R as pitch, and V_R as yaw. AP, ML, and V are used to represent axes in this report (i.e., instead of z, y, and x axes) to ensure consistent labeling between studies. While rotation in degrees is not directly measured from an IMU, the integral of the gyroscope data can be used [6]. Similarly, displacement and velocity can be derived from acceleration parameters [7,8].
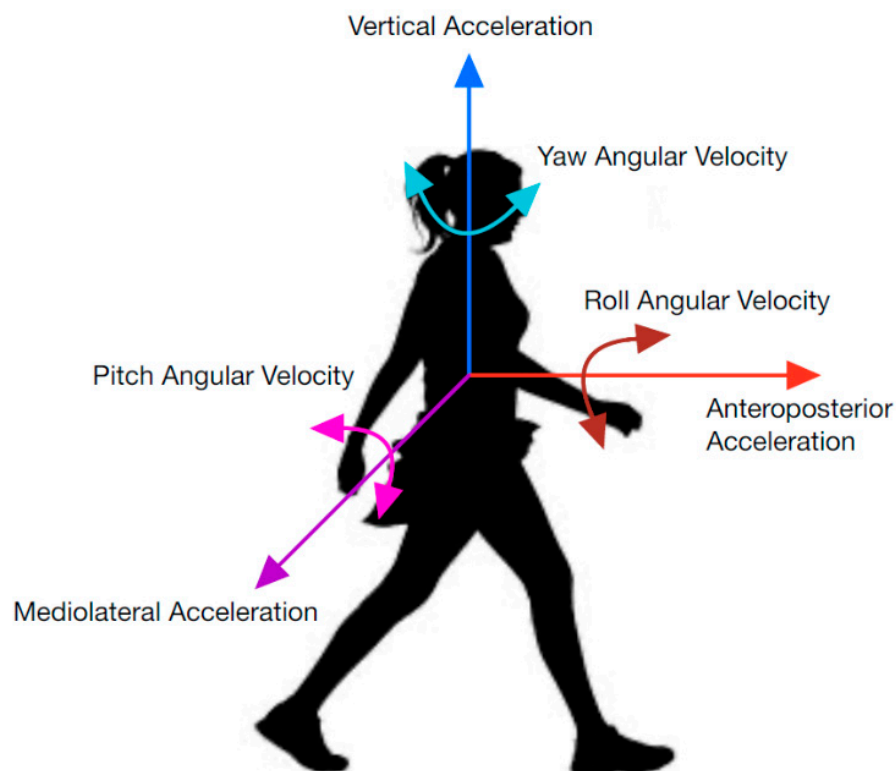

Figure 1. Parametric directions used in inertial data.

In 2015, a methodological review discussed technologies available to assist in TUG use [9]. The review covered a wide variety of sensor and video capture options available for TUG; however, the details of how inertial parameters were used to construct an algorithm for subtask segmentation was out of the review's scope [9]. Another systematic review examined stand up and sit down tasks, but was not associated with the TUG or L Tests [10].

In this paper, we present a scoping review that examines the range of segmentation techniques using inertial parameters, identifies commonly used parameters, and summarizes results between algorithms.

2. Materials and Methods

A scoping review of the literature was conducted using PubMed, Scopus, and Google Scholar. This review follows the PRISMA guidelines for scoping reviews [11]. The searches contained the following terms: "Timed Up and Go AND 'Segment' AND ('turn' OR 'stand'

OR 'sit') AND 'IMU'"; "L Test AND 'Segment' AND ('turn' OR 'stand' OR 'sit') AND 'IMU' AND 'Functional Mobility'". Articles were identified that described subtask segmentation methods for TUG and L Test using only inertial sensor data. Only articles in English were considered. If more than one paper detailed the same segmentation method, then both articles would be referenced for information regarding the segmentation method, and whichever was published first would be included in the identified records list. The articles included were published between the years 2008 and 2022. Separate searches were done for the TUG and L Test. Review of the articles was completed by one reviewer. Titles, abstracts, and then contents were scanned for inclusion criteria. Reference lists of chosen articles were observed for further possible articles. To ensure relevance, a spreadsheet was constructed to summarize the characteristics of each algorithm, including parameters used, population, and subtasks identified. Articles were excluded due to supporting data not from an inertial sensor, no description of the algorithm used, or not segmenting the TUG or L Test. The study selection process is shown in Figure 2.
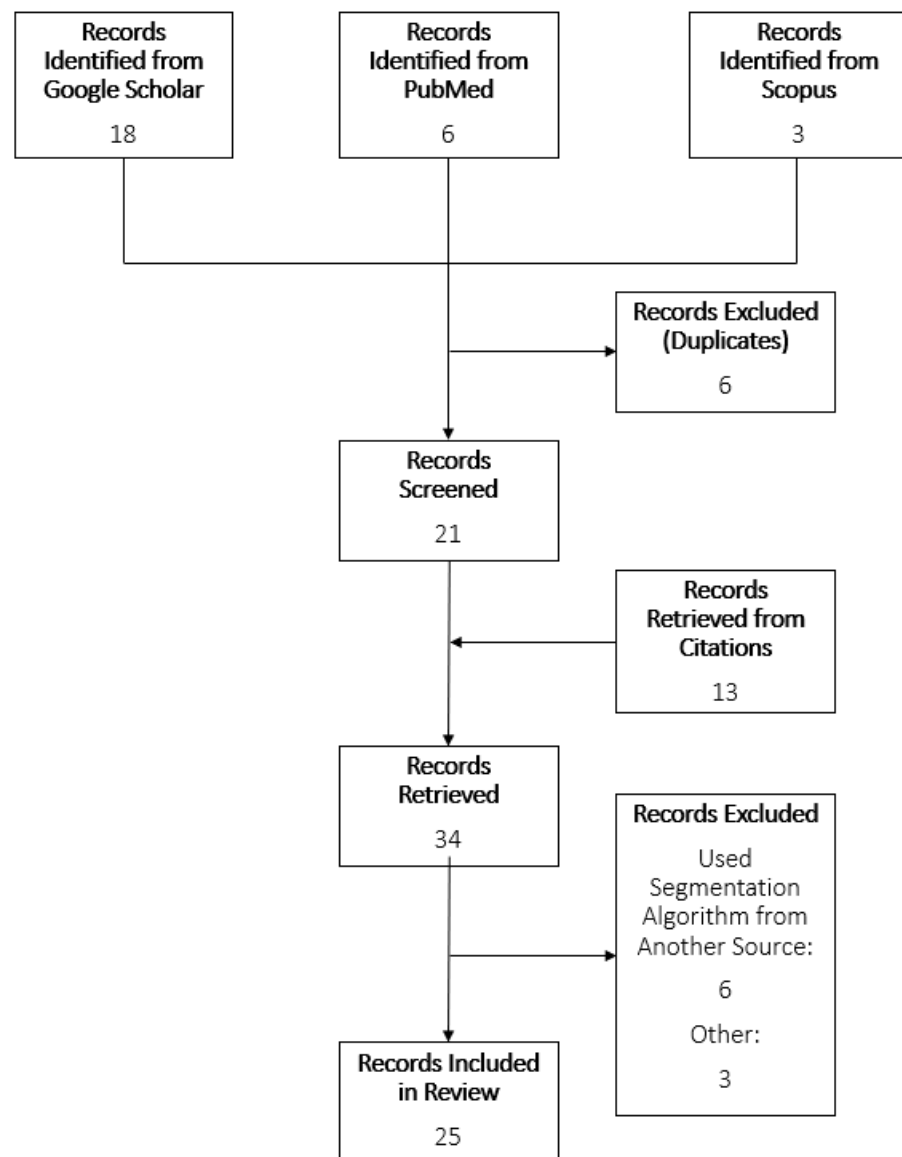

Figure 2. Flowchart of study selection process.

3. Results

Table 1 provides a summary of the articles described in this review. Table 2 provides a summary of results for each article.

Table 1. Summary of articles in this review, sorted by year published.

Source	Test	Sample	# Sensors: Type	Sensor Location	Type
Higashi Y, et al. [3]	TUG	10 healthy, 20 hemiplegic	2: Gyrocube	Waist, leg	Rule-Based
Weiss A, et al. [12]	TUG	15 healthy adults, 17 PD	1: Analog Devices ADXL330	Lower back	Rule-Based
Salarian A, et al. [13]	TUG	12 healthy, 12 PD	8: 3D acceletometer, three 2D Gyroscopes, four single-axis gyroscopes	Sternum, forearms, shanks, thighs,	Rule-Based, Machine Learning
Greene B, et al. [14]	TUG	142 healthy adults; 207 fallers	2: SHIMMER kinematic sensors	Shins	Rule-Based
Jallon P, et al. [15]	TUG	19 healthy adults	1: triaxial accelerometer & magnetometer	Chest	Machine Learning
Adame M, et al. [16]	TUG	10 healthy adults, 20 PD	1: DynaPort Hybrid,	Lower back	Rule-Based
Milosevic, et al. [17]	TUG	4 healthy adults, 3 PD	1: smartphone	Sternum	Rule-Based
Zakaria N, et al. [18]	TUG	38 elderly	1: Accelerometer—MMA7260Q Freescale semiconductor inc + 1D gyroscopes—Murata, Kyoto, Japan, ENC-03R and XV-3500CB, Epson Toyocom, Miyazaki Epson	Lower back	Rule-Based
Nguyen H, et al. [8]	TUG	16 elderly	17: Animazoo IGS-180 motion capture suit	Head, trunk, hip, scapula, upper arm, forearm, hand, thigh, shin, ankle, foot	Rule-Based
Silva K et al. [6]	TUG	18 elderly	1: smartphone	Pocket or waist or leg	Rule-Based
Vervoort D, et al. [19]	TUG	59 healthy adults	1: DynaPort Hybrid	Lower back	Rule-Based
Beyea J, et al. [20]	TUG	10–11 healthy adults	1: 9-axis Microstrain 3DM-GX1 IMU	Upper back	Rule-Based
Negrini S, et al. [21]	TUG	80 healthy adults (some elderly)	1: G-Sensor device	Lower back	Rule-Based
Nguyen H, et al. [22]	TUG	12 early stage PD	17: Animazoo IGS-180 motion capture suit	Covering each body segment	Rule-Based
Hellmers S et al. [4]	TUG	148 elderly, 39 healthy young adults	1: Bosch BMA180 + STMicroelectronics L3GD20H + magnetometer + barometer	Lower back	Machine Learning
Yahalom G, et al. [23]	TUG	25 healthy adults, 25 NPH, 15 PD	1: smartphone	Sternum	Rule-Based
Miller Koop M, et al. [24]	TUG	30 PD	1: iPad	Lower back	Rule-Based
De Luca V, et al. [25]	TUG	20 healthy adults	6: one actibelt + one BioStampRC + four Shimmero	Waist, chest, lower back, ankles	Machine Learning
Witchel H, et al. [26]	TUG	23 healthy, 17 MS	3: x-IMU by x-io	Lateral lower left thigh, lower right thigh, lower back	Rule-Based
Pew C, et al. [27]	L Test	5 amputees	1: iPecs 6-axis load cell	Shank	Machine Learning
Ortega-Bastidas et al. [28]	TUG	25 healthy adults, 12 elderly	1: three-axis accelerometer + three-axis gyroscope + three-axis magnetometer	Lower back	Rule-Based
Hsieh CY, et al. [29]	TUG	5 healthy, 5 OA	3: OPAL sensor	Waist, Thighs, Wrist	Rule-Based
Hsieh CY, et al. [30]	TUG	26 severe knee OA	6: OPAL sensor	Chest, lower back, thighs, shanks	Machine Learning
Abdollah V, et al. [7]	TUG	12 healthy adults	1: Phybrata sensor	Head	Rule-Based
Matey-Sanz M, et al. [31]	TUG	5 healthy adults (testing), 1 healthy adult (training)	1: Smartwatch	Wrist	Machine Learning

PD = Parkinson's Disease; NPH = normal pressure hydrocephalus; MS = multiple sclerosis; OA = osteoarthritis; Rule-Based = a classification model designed and changed by human intervention; Machine Learning = a classification model produced by a machine using a pre-determined algorithm.

Table 2. Results and ground truth comparison methods for each article.

Source	Ground Truth	Stand Up	Sit Down	Turn	Walk	Overall Result
Higashi Y, et al. [3]	Video	r = 0.96	r = 0.74	r = 0.81–0.90	r = 0.81–0.87	
Weiss A, et al. [12]	Stopwatch					
Salarian A, et al. [13]	Video					
Greene B, et al. [14]	Stopwatch			ρ = 0.83	ρ = 0.89–0.90	
Jallon P, et al. [15]						85% detection rate
Adame M, et al. [16]	Observational Labeling	Max Abs Err Dev = 0.22–0.90 s (healthy) Max Abs Err Dev = 0.19–0.81 s (early PD) Max Abs Err Dev = 0.52–0.74 s (long time PD)	Max Abs Err Dev = 0.70–0.95 s (healthy) Max Abs Err Dev = 0.49–0.81 s (early PD) Max Abs Err Dev = 0.73–0.85 s (long time PD)	Max Abs Err Dev = 0.49–1.43 s (healthy) Max Abs Err Dev = 0.41–0.94 s (early PD) Max Abs Err Dev = 0.44–0.71 s (long time PD)		
Milosevic, et al. [17]						
Zakaria N, et al. [18]						
Nguyen H, et al. [8]	Motion Capture	Sn = 100% Sp = 100% Time Diff = 0.03 s ± 0.03 s	Sn = 100% Sp = 100% Time Diff = 0.06 s ± 0.07 s	Sn = 100% Sp = 100% Time Diff = 0.08 ± 0.10 s to 0.18 ± 0.17 s	Sn = 100% Sp = 100% Time Diff = 0.06 ± 0.07 s to 0.18 ± 0.17 s	Sn = 100% Sp = 100%
Silva K et al. [6]	Video					
Vervoort D, et al. [19]						
Beyea J, et al. [20]	Motion Capture	RES = −0.01 ± 0.23 to 0.28 ± 0.33 s	RES = 0.09 ± 0.24 to 0.38 ± 0.30 s	RES = 0 ± 0.30 to 0.27 ± 0.14 s	RES = 0.00 ± 0.23 to −0.30 ± 0.21 s	Abs Err < ±0.25 s, Max Expected Err = 0.34 s
Negrini S, et al. [21]	Motion Capture					RMS dev = 0.35 ± 0.20 s Avg Bias = 0.18 ± 0.21 s [−0.37 ± 0.25 s; 0.74 ± 0.38 s]

Table 2. *Cont.*

Source	Ground Truth	Stand Up	Sit Down	Turn	Walk	Overall Result
Nguyen H, et al. [22]	Motion Capture	Sn = 100% Sp = 100% Time Diff = 0.26 ± 0.29 s	Sn = 100% Sp = 100% Time Diff = 0.19 ± 0.13 s	Sn = 100% Sp = 100% Time Diff = 0.26 ± 0.18 to 0.61 ± 0.18 s	Sn = 100% Sp = 100% Time Diff = 0.26 ± 0.18 to 0.46 ± 0.16 s	Sn = 100% Sp = 100% Time Diff = 0.35 ± 0.16 s
Hellmers S et al. [4]	Stopwatch, Instrumented Chair	Re = 0.84 Pr = 0.66 Acc = 0.99 F1-score = 0.74	Re = 0.94 Pr = 0.56 Acc = 0.99 F1-score = 0.7	Re = 0.78 Pr = 0.83 Acc = 0.99 F1-score = 0.81	Re = 0.98 Pr = 0.98 Acc = 0.98 F1-score = 0.97	
Yahalom G, et al. [23]	Stopwatch					
Miller Koop M, et al. [24]	Observational Labeling			$r^2 = 0.99$		
De Luca V, et al. [25]	Stopwatch	Acc = 87%		Acc = 67%	Acc = 72%	Ω_{shA} = 76%
Witchel H, et al. [26]						
Pew C, et al. [27]	Motion Capture			Acc = 96% (SVM) Acc = 93% (kNN) Acc = 91% (ensemble)	Acc = 85% (SVM) Acc = 82% (kNN) Acc = 97% (ensemble)	
Ortega-Bastidas et al. [28]	Video	Avg Err = −0.08 ± 0.15 s (healthy) Avg Err = 0.05 ± 0.30 s (older) r = 0.81	Avg Err = −0.01 ± 0.19 s (healthy) Avg Err = 0.01 ± 0.29 s (older) r = 0.97–0.98	Avg Err = −0.08 ± 0.11 to −0.19 ± 0.21 s (healthy) Avg Err = −0.01 ± 0.27 to 0.14 ± 0.62 s (older) r = 0.95	Avg Err = 0.17 ± 0.11 to 0.42 ± 0.2 s (healthy) Avg Err = 0.19 ± 0.78 to 0.23 ± 0.61 s (older) r = 0.89–0.99	
Hsieh CY, et al. [29]	Video	Re = 98–100% (healthy) Re = 99% (OA) Pr = 82–86% (healthy) Pr = 81–83% (OA)	Re = 94–98% (healthy) Re = 82–86% (OA) Pr = 82–88% (healthy) Pr = 94–95% (OA)	Re = 79–97% (healthy) Re = 92–99% (OA) Pr = 96–99% (healthy) Pr = 94–97% (OA)	Re = 96–99% (healthy) Re = 94–99% (OA) Pr = 97–99% (healthy) Pr = 98–100% (OA)	Re = 94–95% (healthy) Re = 95% (OA) Pr = 94–95% (healthy) Pr = 94–94% (OA) Acc = 95% (healthy) Acc = 95% (OA)

Table 2. *Cont.*

Source	Ground Truth	Stand Up	Sit Down	Turn	Walk	Overall Result
Hsieh CY, et al. [30]	Video	Sn = 85% Pr = 91%	Sn = 91% Pr = 92%	Sn = 92–89% Pr = 86–94%	Sn = 98% Pr = 94–96%	AdaBoost with 96 sample window size: Sn = 91% Pr = 93% Acc = 94%
Abdollah V, et al. [7]	Motion Capture	Acc = 95% (mastoid) Acc = 93% (sternum) Sn = 90% (mastoid) Sn = 90% (sternum) Sp = 100% (mastoid) Sp = 90% (sternum)	Acc = 98% (mastoid) Acc = 99% (sternum) Sn = 96% (mastoid) Sn = 98% (sternum) Sp = 100% (mastoid) Sp = 100% (sternum)			
Matey-Sanz M, et al. [31]	Video	Mean of Diffs = 0.11 s [−0.56, 0.78] RMSE[18] = 0.38 s	Mean of Diffs = 0.0094 s [−0.62, 0.64] RMSE = 0.32 s	Mean of Diffs = 0.54 s [−0.37, 1.4] RMSE = 0.71 s	Mean of Diffs = −0.25 s [−1.2, 0.74] RMSE = 0.56 s	

ρ = correlation between manual TUG time; Max Abs Err Dev = maximum absolute error deviation; PD = Parkinson's Disease; Sn = sensitivity; Sp = specificity; Time Diff = difference in time; RES = relative error; Abs Err = Absolute Error; Max Expected Err = maximum expected error; RMS dev = root mean squared deviation; Avg Bias = average bias; [1]Re = recall; Pr = precision; Acc = accuracy; Ω_{shA} = percentage of correctly classified participants using the waist accelerometer; Avg Err = average error; Mean of Diffs = mean of differences; RMSE = root mean squared error.

Abdollah et al. [7] had conflicting values in their report for the sensitivity and specificity of stand up in the mastoid sensor. The values from the table in their report were used [7]. The algorithm published by Beyea et al. [20] was also published in a thesis [32].

Seventy-two percent of studies used data from healthy adult participants. The next most common samples were Parkinson's Disease (present in 32% of studies), and elderly (present in 24% of studies). To obtain a higher sensitivity in their machine learning algorithm for stand up, turn, and sit down subtasks, one study with elderly participants recruited an additional 39 healthy adults [4].

Half of the studies used a single IMU placed on the participant's lower back [4,12,16,18,19,21,24,28], three chest/sternum [15,17,23], and four at other locations [6,7,20,31]. Lower back placement approximates whole body center of mass [24]. Beyea [32] originally attempted accelerometer placement on the chest and right shoulder, as well as a goniometer on the knee; however, they later used only IMU signals due to the accelerometers being deemed inadequate and the goniometer not providing clearer data than the IMU. Negrini et al. [21] used three sensors, and analyzed each individually.

Pre-processing calibration was mentioned in multiple studies, with a directional cosine matrix used to calibrate iPad inertial measurements to ensure that gravity was acting in only the vertical direction [24], or, static data from each participant used to orient the sensor [2,16]. Wiess et al. [12] also mentioned a calibration algorithm; however, filtering methods were not provided.

Zero-lag low-pass Butterworth filters were sometimes used for pre-processing, with varying orders and cut-off frequency levels [3,14,20,24,26]. One study used a cut-off frequency equal to the walking cadence [3]. Some studies used low-pass filters with a variety of frequencies [4,13,18,23,31]. A moving average filter was also used [16,27,29,30]. Multi sensor studies sometimes used a variety of cut-off frequencies for filtering [26,30]. Nguyen et al. [8,22] used bandpass filtering. Some studies also chose to normalize data before identifying tasks [20–22,22,28,31].

Algorithm types were divided into rule-based and machine learning, with 76% of studies using rule-based algorithms and 28% using machine learning algorithms. One study used a combination of rule-based and machine learning techniques [13] and another study used a dynamic time warping technique [16]. Both rule-based and machine learning approaches implemented normalization, filtering, and feature calculations during pre-processing; however, no preprocessing trends were apparent. Some studies included a subtask-based temporal order in their algorithm to ensure that any artifact or short discrepancy in the signal would not lead to misclassification [4,30]. Calculation of other features from inertial parameters, such as mean, standard deviation, and skewness were sometimes reported [30].

3.1. Stand Up and Sit Down

Using a rule-based algorithm, the stand up subtask was often identified at the time that the participant began to lean forward (i.e., beginning of the first peak caused by a change in mediolateral angular velocity (ML_ω)) [16,17,20,28]. A similar processes could be used with mediolateral rotation (ML_R) peak [7,13,18], and/or anteroposterior acceleration (AP_a) [12,19,21]. Beyea et al. [20] used ML_ω, AP_a, and vertical acceleration (V_a) and Silva et al. [6] used ML_R and identified consecutive 3° changes to be part of the stand up task. Another method for labeling stand up movements identified a threshold value for $ML_{R,sternum}$, then a logistic regression model was used to give a probability for each candidate [13].

For multi sensor analysis, one study [8] used 17 sensors but only used data from the hip, knee, head, and trunk, with the stand up subtask using $AP_{a,trunk}$ and $ML_{\omega,hip}$. An updated study [22] took measurements from the hip, thigh, and trunk, and used $V_{a,trunk}$, $AP_{a,thigh}$, and $ML_{r,hip}$ for sit-to-stand. In other multi sensor studies, $ML_{\omega,trunk}$ [29] and $ML_{\omega,waist}$ [3] were used. Some multi sensor studies averaged all ML_ω sensors [14,26].

Most studies identified stand up and sit down tasks using similar methods, and labeled respective tasks based on temporal order [3,7,8,12,13,18–21,26,28]. Hsieh et al. [29] identified the sit down task as when $V_{a,thigh}$ reached a minimum acceleration.

3.2. Walking

The beginning of walking could be identified when the participant stopped standing up and began to walk forward, indicated by the end of the first peak in V_a [32]. AP_a was also used to identify the beginning of walking, with either a threshold [24] or peak identification method [19,21]. Alternatively, walking was considered a "null signal", where only stand up, sit down, and turn tasks were identified and a temporal order identified movement between these tasks as walking [7,16,18,20,26,28]. Nguyen et al. [22] identified walking using $AP_{a,shin}$ and $ML_{r,hip}$, and achieved a sensitivity and specificity of 100% and a maximum difference in time compared to their ground truth values of 612 ± 175 ms.

3.3. Turns

Turns were sometimes identified using vertical angular velocity, V_ω, with the start of a turn at the beginning of a big peak in V_ω and the end of the turn at the corresponding peak end [20–22,32] One study that used this method attained a relative error of 0.00 ± 0.30 s to 0.27 ± 0.14 s [20]. V_ω was also used with trunk and head sensors [8] or the waist sensor [3,29].

Another turn identification method used a threshold value (or observed for increase/decrease [18]) for vertical rotation (V_R) [6,18,19,24,28,31]. Additionally, using V_R, a sliding window technique was attempted to identify the end of the turn, and had a maximum average error of -0.19 ± -0.21 s [28].

A more complex method used V_R with a least squares optimization algorithm to identify turns and reported $r^2 = 0.99$ [24]. Greene et al., identified that $ML_{\omega,shin}$ was lower during turning than walking, and used this relationship to identify turns ($\rho = 0.83$) [14].

3.4. Machine Learning

In a study using machine learning techniques, five-fold cross-validation was used, with one fifth of data kept for testing, then repeated five times so that each fold was used for validation once [4]. As the only hierarchical approach to machine learning subtask segmentation identified in this review, the data were classified into "Static", "Dynamic", and "Transition" categories using Boosted Decision trees. Then, the data were further classified into subtasks using Multilayer Perceptions [4]. Jallon et al. [15] proposed a graph-based approach where Bayesian, Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) techniques were analyzed. The graph enforced Bayesian classifier was the most successful, accomplishing subtask detection at a rate of approximately 85%.

Matey-Sanz et al. [31] proposed a training algorithm using a Multilayer Perceptron model created using TensorFlow Lite, and De Luca et al. [25] proposed an algorithm using k-means clustering. Hsieh et al. [30] compared SVM, K-Nearest Neighbor (kNN), Naïve Bayesian (NB), Decision Tree (DT), and Adaptive Boosting (AdaBoost) machine learning algorithms, and ultimately found that AdaBoost achieved the best results with 90.62% sensitivity, 93.03% precision, and 94.29% accuracy.

The one study segmenting L Test data compared SVM, kNN, and Ensemble techniques [27]. SVM had the highest accuracy for turns at 96%, while Ensemble had the highest accuracy for walking straight at 97% [27].

Seven studies did not compare their algorithm, or individual subtasks, to a ground truth measurement [6,12,17–19,24,26]. The results of these studies, therefore, cannot be analyzed.

4. Discussion

In this paper, 25 studies were identified that presented IMU-based methods of subtask segmentation for the TUG and L Test. The methods outlined in the literature used mainly threshold or maximum identification techniques; however, seven used machine learning

techniques. For rule-based algorithms, ML_ω (or its derivative, ML_R) was more often used for stand up and sit down tasks (which most commonly were classified using the same algorithm and individually identified using temporal labeling). Similarly, V_ω and its derivative V_R was used for turning classification in all but one of the rule-based algorithms that considered turning in their segmentation. In eight studies, the gait subtask was identified as a null parameter, therefore being classified as the time between appropriate tasks, such as stand up and turn. While this may often be acceptable for able-bodied individuals, other populations may pause, for example, between the stand up and walk tasks, which would be a moment in which gait should not be analyzed as the person would not yet be walking.

Machine Learning algorithms often had lower accuracies than rule-based methods [4,7,25,27,29,30]. Boosted Decision Trees achieved the highest overall accuracy by a machine learning approach (96.55% in elderly individuals), which was similar or above some rule-based methods [4,7,29]. To assess the viability of these models, statistical results from Table 2 can be compared to literature on the reliability of TUG measures. In a study by Botolfsen et al. [33], it was shown that for patients with fibromyalgia, the standard error of measurement for the TUG timed with a stop watch was 0.231 s.

In three studies, a stopwatch was used as the ground truth measurement for task segmentation [4,12,23]. This is not recommended, since a stopwatch introduces variability and decreases precision in comparison to other methods [32]. If stopwatch data are used to train a model, then the resulting model cannot provide outcomes better than the stopwatch approach, which relies on a person properly marking transitions in real time during the test. While Hellmers et al. [4] wanted to work toward an unsupervised approach (where a stopwatch would not be feasible), different ground truth comparison methods could provide machine learning training data that are not influenced by human reaction time. Video-recorded or motion-captured data are often used as ground truth records in TUG segmentation, providing higher accuracy of measurement than the stopwatch [7,21,32]. Beyea [32] used Vicon rather than video recordings due to the ease of processing large quantities of data.

Beyea et al. [20] noted that the calibration and alignment could affect threshold segmentation methods if the data were not first normalized. Salarian et al. [13,34] also noted that gait and transitions could cause noise, to which a threshold-based algorithm may be sensitive.

Limitations and Criticisms

Numerous studies collected data from fewer than 15 participants, which is insufficient to avoid type II error [7,17,20,29,31,32]. The machine learning algorithm described by Matey-Sanz et al. [31] used training data from a single participant, and consequently had the highest recorded root mean squared error (in seconds). A larger sample size would allow machine learning algorithms to train to adjust to different walking patterns and speeds, and also help identify general issues with rule-based systems.

There were several studies in which the proposed algorithm was not clearly explained. An assumed algorithm was attained from a figure in the study conducted by Yahalom et al. [23], and there were no further connecting papers related to their process of automatic segmentation. In a study by Koop et al. [24], part of the algorithm was described as "using the linear acceleration and angular velocity in the (AP, ML, and V) direction in conjunction with previously reported methods adjusted with data specific parameters". However, no previously reported methods were cited. If a study did not clearly describe their subtask segmentation but discussed increase/decrease in parameters throughout their methodology, then threshold values were assumed to be used [18,21]. In [32], it is not clear which axes are used for vertical and anteroposterior directions.

Instead of a fixed walking distance, Silva et al. [6] used a timed approach, where the participant was given 15 seconds to stand up and walk forward, and then 15 seconds to

turn, walk back, and sit down. This resulted in lost data, and therefore only the first stand up, walk, and turn tasks were identified.

5. Conclusions

A variety of reporting methods for TUG subtask segmentation were found in the literature, often with a paucity of implementation details given. Future investigations should consider individual algorithms for standing, sitting, walking, and turning. This could allow for each subtask identification to be improved. Other research that augments clinical mobility tests using inertial sensors has found that applying AI models can provide additional information not typically related to the clinical test, such as fall risk from six-minute walk test data [35]. A segmented L Test could provide similar AI enhanced knowledge; for example, balance confidence, task specific performance, or improved fall risk classification. The L Test also requires full analysis and automated segmentation, to ensure that left and right turns are appropriately classified.

Furthermore, using larger data sets for comparing different algorithms is necessary for in-depth comparisons. In cases where data collection may be difficult, data augmentation should be researched, depending on the number of samples collected. A study conducted on biomechanics, "CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data" [36] stated that although data augmentation could be used to improve robustness, it cannot account for variation in movement patterns. Therefore, the recommended number of participants for biomechanics studies (up to 25 in some cases [37]) would still be required. Further research on data augmentation is needed to explore the limits and benefits for this application.

To further understand the performance between rule-based and machine learning algorithms, an investigation should also be performed in the improvements that machine learning has been shown to have compared to rule-based algorithms, where data have artifacts or noise.

Author Contributions: Conceptualization, A.L.M.F., N.B. and E.D.L.; methodology, A.L.M.F.; formal analysis, A.L.M.F.; writing—original draft preparation, A.L.M.F.; writing—review and editing, A.L.M.F., N.B. and E.D.L.; supervision, N.B. and E.D.L.; funding acquisition, N.B. and E.D.L. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sebastião, E.; Sandroff, B.M.; Learmonth, Y.C.; Motl, R.W. Validity of the Timed Up and Go Test as a Measure of Functional Mobility in Persons With Multiple Sclerosis. *Arch. Phys. Med. Rehabil.* **2016**, *97*, 1072–1077. [CrossRef]
2. Haas, B.; Clarke, E.; Elver, L.; Gowman, E.; Mortimer, E.; Byrd, E. The reliability and validity of the L-test in people with Parkinson's disease. *Physiotherapy* **2019**, *105*, 84–89. [CrossRef]
3. Higashi, Y.; Yamakoshi, K.; Fujimoto, T.; Sekine, M.; Tamura, T. Quantitative evaluation of movement using the timed up-and-go test. *IEEE Eng. Med. Biol. Mag.* **2008**, *27*, 38–46. [CrossRef]
4. Hellmers, S.; Izadpanah, B.; Dasenbrock, L.; Diekmann, R.; Bauer, J.M.; Hein, A.; Fudickar, S. Towards an Automated Unsupervised Mobility Assessment for Older People Based on Inertial TUG Measurements. *Sensors* **2018**, *18*, E3310. [CrossRef]
5. Li, T.; Chen, J.; Hu, C.; Ma, Y.; Wu, Z.; Wan, W.; Huang, Y.; Jia, F.; Gong, C.; Wan, S.; et al. Automatic Timed Up-and-Go Sub-Task Segmentation for Parkinson's Disease Patients Using Video-Based Activity Classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 2189–2199. [CrossRef]
6. Silva, J.; Sousa, I. Instrumented timed up and go: Fall risk assessment based on inertial wearable sensors. In Proceedings of the 2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Benevento, Italy, 15–18 May 2016; pp. 1–6.
7. Abdollah, V.; Dief, T.N.; Ralston, J.; Ho, C.; Rouhani, H. Investigating the validity of a single tri-axial accelerometer mounted on the head for monitoring the activities of daily living and the timed-up and go test. *Gait Posture* **2021**, *90*, 137–140. [CrossRef]

8. Nguyen, H.P.; Ayachi, F.; Lavigne–Pelletier, C.; Blamoutier, M.; Rahimi, F.; Boissy, P.; Jog, M.; Duval, C. Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults using multiple inertial sensors. *J. Neuro Eng. Rehabil.* **2015**, *12*, 36. [CrossRef]

9. Sprint, G.; Cook, D.J.; Weeks, D.L. Toward Automating Clinical Assessments: A Survey of the Timed Up and Go. *IEEE Rev. Biomed. Eng.* **2015**, *8*, 64–77. [CrossRef]

10. Millor, N.; Lecumberri, P.; Gomez, M.; Martínez-Ramirez, A.; Izquierdo, M. Kinematic Parameters to Evaluate Functional Performance of Sit-to-Stand and Stand-to-Sit Transitions Using Motion Sensor Devices: A Systematic Review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2014**, *22*, 926–936. [CrossRef]

11. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef]

12. Weiss, A.; Herman, T.; Plotnik, M.; Brozgol, M.; Maidan, I.; Giladi, N.; Gurevich, T.; Hausdorff, J.M. Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease? *Med. Eng. Phys.* **2010**, *32*, 119–125. [CrossRef]

13. Salarian, A.; Horak, F.B.; Zampieri, C.; Carlson-Kuhta, P.; Nutt, J.G.; Aminian, K. iTUG, a sensitive and reliable measure of mobility. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2010**, *18*, 303–310. [CrossRef]

14. Greene, B.R.; O'Donovan, A.; Romero-Ortuno, R.; Cogan, L.; Scanaill, C.N.; Kenny, R.A. Quantitative Falls Risk Assessment Using the Timed Up and Go Test. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2918–2926. [CrossRef]

15. Jallon, P.; Dupre, B.; Antonakios, M. A graph based method for timed up & go test qualification using inertial sensors. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 689–692.

16. Adame, M.R.; Al-Jawad, A.; Romanovas, M.; Hobert, M.A.; Maetzler, W.; Möller, K.; Manoli, Y. TUG Test Instrumentation for Parkinson's disease patients using Inertial Sensors and Dynamic Time Warping. *Biomed. Eng. Biomed. Tech.* **2012**, *57*, 1071–1074. [CrossRef]

17. Milosevic, M.; Jovanov, E.; Milenković, A. Quantifying Timed-Up-and-Go test: A smartphone implementation. In Proceedings of the 2013 IEEE International Conference on Body Sensor Networks, Cambridge, MA, USA, 6–9 May 2013; pp. 1–6.

18. Zakaria, N.A.; Kuwae, Y.; Tamura, T.; Minato, K.; Kanaya, S. Quantitative analysis of fall risk using TUG test. *Comput. Methods Biomech. Biomed. Engin.* **2013**, *18*, 426–437. [CrossRef]

19. Vervoort, D.; Vuillerme, N.; Kosse, N.; Hortobágyi, T.; Lamoth, C.J.C. Multivariate Analyses and Classification of Inertial Sensor Data to Identify Aging Effects on the Timed-Up-and-Go Test. *PLoS ONE* **2016**, *11*, e0155984. [CrossRef]

20. Beyea, J.; McGibbon, C.A.; Sexton, A.; Noble, J.; O'Connell, C. Convergent Validity of a Wearable Sensor System for Measuring Sub-Task Performance during the Timed Up-and-Go Test. *Sensors* **2017**, *17*, 934. [CrossRef]

21. Negrini, S.; Serpelloni, M.; Amici, C.; Gobbo, M.; Silvestro, C.; Buraschi, R.; Borboni, A.; Crovato, D.; Lopomo, N.F. Use of Wearable Inertial Sensor in the Assessment of Timed-Up-and-Go Test: Influence of Device Placement on Temporal Variable Estimation. In Proceedings of the Wireless Mobile Communication and Healthcare; Perego, P., Andreoni, G., Rizzo, G., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 310–317.

22. Nguyen, H.; Lebel, K.; Boissy, P.; Bogard, S.; Goubault, E.; Duval, C. Auto detection and segmentation of daily living activities during a Timed Up and Go task in people with Parkinson's disease using multiple inertial sensors. *J. Neuro Eng. Rehabil.* **2017**, *14*, 26. [CrossRef]

23. Yahalom, G.; Yekutieli, Z.; Israeli-Korn, S.; Elincx-Benizri, S.; Livneh, V.; Fay-Karmon, T.; Rubel, Y.; Tchelet, K.; Zauberman, J.; Hassin-Baer, S. AppTUG-A Smartphone Application of Instrumented 'Timed Up and Go' for Neurological Disorders. *EC Neurol.* **2018**, *10*, 689–695.

24. Miller Koop, M.; Ozinga, S.J.; Rosenfeldt, A.B.; Alberts, J.L. Quantifying turning behavior and gait in Parkinson's disease using mobile technology. *IBRO Rep.* **2018**, *5*, 10–16. [CrossRef]

25. De Luca, V.; Muaremi, A.; Giggins, O.M.; Walsh, L.; Clay, I. *Towards Fully Instrumented and Automated Assessment of Motor Function Tests*; IEEE: New York, NY, USA, 2018; pp. 83–87.

26. Witchel, H.J.; Oberndorfer, C.; Needham, R.; Healy, A.; Westling, C.E.I.; Guppy, J.H.; Bush, J.; Barth, J.; Herberz, C.; Roggen, D.; et al. Thigh-Derived Inertial Sensor Metrics to Assess the Sit-to-Stand and Stand-to-Sit Transitions in the Timed Up and Go (TUG) Task for Quantifying Mobility Impairment in Multiple Sclerosis. *Front. Neurol.* **2018**, *9*, 684. [CrossRef]

27. Pew, C.; Klute, G.K. Turn Intent Detection For Control of a Lower Limb Prosthesis. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 789–796. [CrossRef]

28. Ortega-Bastidas, P.; Aqueveque, P.; Gómez, B.; Saavedra, F.; Cano-de-la-Cuerda, R. Use of a Single Wireless IMU for the Segmentation and Automatic Analysis of Activities Performed in the 3-m Timed Up & Go Test. *Sensors* **2019**, *19*, 1647. [CrossRef]

29. Hsieh, C.-Y.; Huang, H.-Y.; Liu, K.-C.; Chen, K.-H.; Hsu, S.J.; Chan, C.-T. *Automatic Subtask Segmentation Approach of the Timed Up and Go Test for Mobility Assessment System Using Wearable Sensors*; IEEE: New York, NY, USA, 2019.

30. Hsieh, C.-Y.; Huang, H.-Y.; Liu, K.-C.; Chen, K.-H.; Hsu, S.J.-P.; Chan, C.-T. Subtask Segmentation of Timed Up and Go Test for Mobility Assessment of Perioperative Total Knee Arthroplasty. *Sensors* **2020**, *20*, 6302. [CrossRef]

31. Matey-Sanz, M.; González-Pérez, A.; Casteleyn, S.; Granell, C. Instrumented Timed Up and Go Test Using Inertial Sensors from Consumer Wearable Devices. In Proceedings of the Artificial Intelligence in Medicine; Michalowski, M., Abidi, S.S.R., Abidi, S., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 144–154.
32. Beyea, J.B.A. Automating the Timed Up and Go Test (Tug Test) with Wearable Sensors. Master's Thesis, University of New Brunswick, Fredericton, NB, Canada, 2017.
33. Botolfsen, P.; Helbostad, J.L.; Moe-nilssen, R.; Wall, J.C. Reliability and concurrent validity of the Expanded Timed Up-and-Go test in older people with impaired mobility. *Physiother. Res. Int.* **2008**, *13*, 94–106. [CrossRef]
34. Salarian, A.; Zampieri, C.; Horak, F.B.; Carlson-Kuhta, P.; Nutt, J.G.; Aminian, K. Analyzing 180° turns using an inertial system reveals early signs of progress in Parkinson's Disease. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; Volume 229, pp. 224–227. [CrossRef]
35. Juneau, P.; Baddour, N.; Burger, H.; Bavec, A.; Lemaire, E.D. Amputee Fall Risk Classification Using Machine Learning and Smartphone Sensor Data from 2-Minute and 6-Minute Walk Tests. *Sensors* **2022**, *22*, 1479. [CrossRef]
36. Dorschky, E.; Nitschke, M.; Martindale, C.F.; van den Bogert, A.J.; Koelewijn, A.D.; Eskofier, B.M. CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data. *Front. Bioeng. Biotechnol.* **2020**, *8*, 604. [CrossRef]
37. Oliveira, A.S.; Pirscoveanu, C.I. Implications of sample size and acquired number of steps to investigate running biomechanics. *Sci. Rep.* **2021**, *11*, 3083. [CrossRef]