AUTOMATIC SUBTASK SEGMENTATION OF THE L TEST

ALEXIS MCCREATH FRANGAKIS

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Applied Science in Biomedical Engineering

Department of Mechanical Engineering
Faculty of Engineering
University of Ottawa

Acknowledgments

I would like to express my deepest gratitude to my mom, dad, sister, and my partner, Heemel, for their unwavering support and encouragement throughout the entirety of this journey.

Abstract

Data gathered from inertial measurement units can be used to design subtask segmentation algorithms that further analyse functional mobility tests. This method has been shown to be effective for the 6-minute walk test, and the timed up-and-go test (TUG), and have produced gait analysis and fall risk models, especially in an elderly population. While further research has recently been published to assess individuals with a more asymmetrical gait, such as lower limb amputees, subtask segmentation of functional mobility tests such as the L Test of Functional Mobility (L Test), which are recommended for lower limb amputees, has not been investigated. With a decreased ceiling effect in comparison to the TUG, and increased turn requirements, the L Test would be able to provide a more in-depth assessment of a patient's mobility.

In this thesis, a rule-based subtask segmentation algorithm was designed and tested with data from both able-bodied individuals and lower limb amputees. The algorithm produced acceptable results for able-bodied individuals (97% accuracy, > 98% specificity, > 74% sensitivity), but had low sensitivity for data from lower limb amputees (93-97% accuracy, 97-99% specificity, 33-60% sensitivity). A machine learning algorithm was then trained on data from both able-bodied and lower limb amputee participants and tested on data from a lower limb amputee population. This algorithm produced improved results for lower limb amputee participants (> 85% accuracy, > 75% sensitivity, > 95% specificity).

This research designed and assessed both a rule-based and a machine learning algorithm for subtask segmentation of the L Test using data collected from both able-bodied and lower limb amputee participants. Overall, this thesis contributes to the progression of movement analysis for lower limb amputees, and to the understanding of motion during an L Test.

Table of Contents

List of Figures

List of Tables

Abbreviations and Definitions

Abs Err	Absolute error
Acc	Accuracy
Accuracy_{AD}	Accuracy within one average step duration
AdaBoost	Adaptive Boosting
AP_A	Linear acceleration on the antero-posterior axis
AP_R	Rotation about the antero-posterior axis
AP_ω	Angular velocity about the antero-posterior axis
Avg Bias	average bias
Avg Err	Average error
DT	Decision Tree
ESPRM	European Society of Physical and Rehabilitation Medicine
IMU	Inertial measurement unit
kNN	k-Nearest Neighbor
LDA	Linear Discriminant analysis
LLA	Lower Limb Amputees
LOO	Leave-one-out
L Test	L Test of Functional Mobility
Max Abs Err Dev	Maximum absolute error deviation
Max Exp Err	Maximum expected error
MDT_{SR}	Magnitude difference threshold for rotation angle at the beginning of the stand-up task
$MDT_{S\omega}$	Magnitude different threshold for angular velocity at the beginning of the stand-up task
MDT_T	Magnitude difference threshold for turns
Mean of diffs	Mean of differences
ML	Machine learning
ML_A	Linear acceleration on the medial-lateral axis
ML_R	Rotation about the medial-lateral axis
ML_ω	Angular velocity about the medial-lateral axis

NB	Naïve Bayesian
OA	Knee osteoarthritis
PD	Parkinson’s Disease
Pr	Precision
Re	Recall
RES	Relative error
RMS dev	Root mean squared deviation
RMSE	Root mean squared error
SD	Standard deviation
SDT_{T}	Standard deviation threshold for turns
SDT_{SR}	Standard deviation threshold for rotation angle at the beginning of the stand-up task
SDT_{StR}	Standard deviation threshold for rotation angle at the end of the stand-up task
$SDT_{S\omega}$	Standard deviation threshold for angular velocity at the beginning of the stand-up task
$SDT_{St\omega}$	Standard deviation threshold for angular velocity at the end of the stand-up task
Sn	Sensitivity
Sp	Specificity
SVM	Support Vector Machines
Time Diff	Difference in time
TOHRC	The Ottawa Hospital Rehabilitation Centre
TUG	Timed Up-And-Go Test
V_{A}	Linear acceleration on the vertical axis
V_{R}	Rotation about the vertical axis
V_{ω}	Angular velocity about the vertical axis
2MWT	2-Minute Walk Test
5STS	5 Times Sit to Stand
6MWT	6-Minute Walk Test
10MWT	10-Minute Walk Test

ρ	Correlation between manual TUG time and algorithm time
Ω_{shA}	Percentage of correctly classified participants using waist accelerometer

1 Introduction

Timed walking tests are valid and reliable tools to assess a person's gait and mobility [1], [2]. The most common tests include the Timed Up & Go (TUG) test, 10-meter walk test (10MWT), and 6-minute walk test (6MWT) [1]. The TUG test was originally designed for clinical observation and quantification of balance in elderly individuals [1], including daily living activities of standing up from a chair, walking three metres, turning around, walking back to the chair, and turning and sitting back down [3]. However, TUG has some drawbacks, especially when used with other populations such as lower limb amputees [3]. Drawbacks include a ceiling effect with increasing mobility level due to the TUG's short distance and that the person may only turn in their preferred direction [3].

The L Test of Functional Mobility (L Test) was proposed to address these drawbacks [3]. This test includes standing up from a chair, walking three metres to a marker, turning 90 degrees, walking 7 metres to a second marker, turning 180 degrees, walking back to the first marker, turning 90 degrees, walking back to the chair, turning 180 degrees, and sitting down [3] (Figure 1.1). Clinicians record the total time to complete the L Test, using a stopwatch, and compare this time to the results from individuals with similar impairments [3]. Additionally, observational qualitative information may be recorded, such as quality of gait or prosthetic use [3].

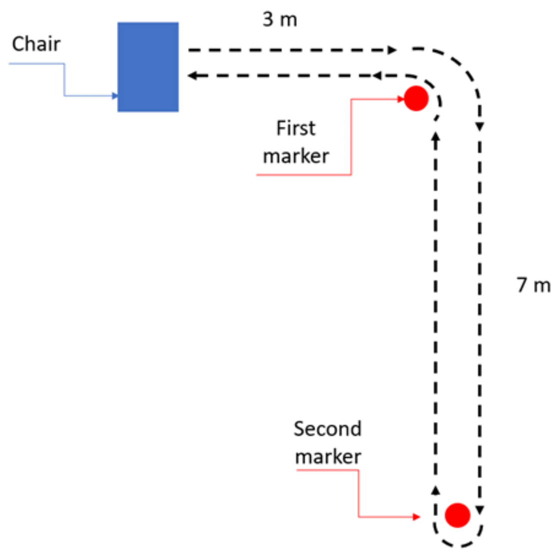


Figure 1.1 L Test of Functional Mobility

With recent advancements in technology, sensors have been used to obtain further performance metrics from functional mobility tests. Subtask (i.e. sit, 180° turn) times have been shown to correlate with fall risk [4], [5] and functional recovery is closely related to stand-up and sit-down times [5]. Additionally, foot strike identification during tests such as the 6MWT is useful for calculating step-based features for further fall risk classification [6]. Segmented data can create a more in-depth and better representation of a patient's movement status, with new information on tasks for which a patient may need more support. A stopwatch is not a good option to obtain subtask timing since manually timing subtasks and foot strikes introduces more variability between clinicians and may increase error, which is already criticized as a major drawback of L Test and TUG performance measures [7]. Additionally, with the clinician focusing on timing subtasks and foot strikes, their ability to gather observational information is reduced. With the possibility of obtaining a richer array of data from functional mobility tests, the use of wearable sensor technology for automatic segmentation has been proposed and shown to be beneficial and lead to clearer treatment pathways [8].

A variety of methods have been proposed to instrument clinical movement-related tests; including, inertial measurement units (IMUs), 3D motion analysis systems, video recording, and instrumented canes [9], [10], [11]. While methods such as 3D motion analysis and video recordings can produce excellent results, they are not ideal for clinical use because the equipment required is often expensive and inaccessible in a clinical setting [12].

Despite the recommendations for the L Test to be used in place of the TUG for many populations [3], research on L Test task segmentation is lacking [12]. Currently, the author has identified no published methods of segmenting the L Test using IMUs [12].

1.1 Rationale

The growing number of elderly people could lead to longer clinic wait times and decreased appointment duration. Automating the segmentation and post processing of functional mobility tests allows for shorter testing, diagnosis, and decision times within a clinical encounter, therefore creating a more efficient clinical experience. Without the design and implementation of a reliable task segmentation technique, the identification of a patient's problem areas (i.e. walking, turning, sitting) relies solely on practitioner observations. Furthermore, high quality evaluation techniques such as motion analysis systems are not accessible or feasible in most clinics due to cost and space restrictions. Hence, the design of a low-cost, simple, unobtrusive system that provides better outcome measures is needed to inform therapy and mitigate fall risk. The ubiquity of smartphones and their powerful sensors motivates their use with clinical tests. This thesis designed and evaluated a new automatic subtask segmentation approach for the L Test using smartphone IMUs. The model initially used data from healthy individuals to baseline an approach. The baseline model was then adjusted and assessed with data from lower limb amputees.

1.2 Objective

The objective of this thesis is to create and evaluate algorithms for automatically identifying L Test subtasks for able-bodied individuals and people with lower limb amputations using smartphone sensor data.

1.3 Thesis Contributions

This thesis contributed to automatic subtask segmentation and smartphone applications for clinical movement evaluation in lower limb amputees. Specific contributions are:

- Development of a viable rule-based algorithm for subtask segmentation of the L Test for able-bodied individuals using smartphone signals. No previous algorithms have been published that automatically segment the full L Test into the stand, walk, turn 90°, turn 180°, and sit tasks. This method and algorithm could be implemented on a smartphone to enhance the quantity and quality of information available to the clinician at the point of patient care.
- Development of a machine learning model for subtask segmentation of the L test for lower limb amputees, using a random forest algorithm. This approach was superior to the rule-based algorithm and provides a basis for further improvements as datasets for model training grow.

1.4 Thesis Outline

This thesis manuscript is divided into 7 chapters.

Chapter 2 is the literature review. This chapter includes a scoping review describing current techniques for subtask segmentation of the TUG and L Test using smartphone IMUs. The contents of this chapter were published in the journal *Information*.

Chapter 3 describes and evaluates an algorithm for subtask segmentation of the L Test for able-bodied individuals. The contents of this chapter were published in the journal *BioMedInformatics*.

Chapter 4 reports on the results of a rule-based algorithm designed for able-bodied individuals, when tested on data from a lower limb amputee population. The contents of this chapter were published in a conference abstract for the *24th Congress of the European Society for Physical and Rehabilitation Medicine*.

Chapter 5 describes and evaluates an algorithm for subtask segmentation of the L Test for lower limb amputees. The contents of this chapter have been submitted for publication.

Chapter 6 presents a thesis summary and suggestions for future work.

2 Subtask Segmentation Methods of the Timed Up and Go and L Test using Inertial Measurement Units – A Scoping Review

This chapter addresses the current state of literature regarding segmentation methods for the TUG and the L Test using smartphone inertial measurement units. It outlines all identified publications that discuss segmentation methods for these tests, and discusses common ground truth identification methods, filtering methods, devices used, demographics, and types of algorithms.

The contents of this chapter were published in the journal ‘Information’:

McCreath Frangakis, A.L.; Lemaire, E.D.; Baddour, N. Subtask Segmentation Methods of the Timed Up and Go Test and L Test Using Inertial Measurement Units—A Scoping Review. Information 2023, 14, 127. https://doi.org/10.3390/info14020127.

2.1 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.

2.2 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 [13]. 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 [13]. 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 [14]. 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 [14]. It also further encompasses everyday movement since the two 90° turns force the patient to turn both left and right, instead of just their preferred way [14].

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 [15]. 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 [15]. 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 [15]. Automated subtask segmentation often uses inertial measurement units (IMU), video recordings, or pressure-sensor data [15], [16], [17]. 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 **Error! Reference source not found.**. They include a nteroposterior 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 [18]. Similarly, displacement and velocity can be derived from acceleration parameters [5], [19].

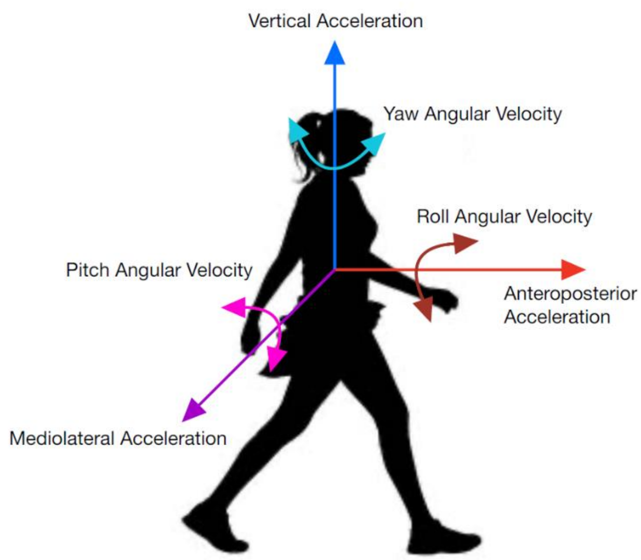


Figure 2.1 Parametric directions used in inertial data

In 2015, a methodological review discussed technologies available to assist in TUG use [20]. The review covered a wide variety of sensor and video capture options available for the TUG; however, the details of how inertial parameters were used to construct an algorithm for subtask segmentation was out of the review's scope [20]. Another systematic review examined stand-up and sit-down tasks, but was not associated with the TUG or L Test [21].

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.

2.3 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 [22]. 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 having supporting data that was not from an inertial sensor, not describing the algorithm used, or not segmenting the TUG or L test. A flowchart of the study selection process is shown in Figure 2.2.

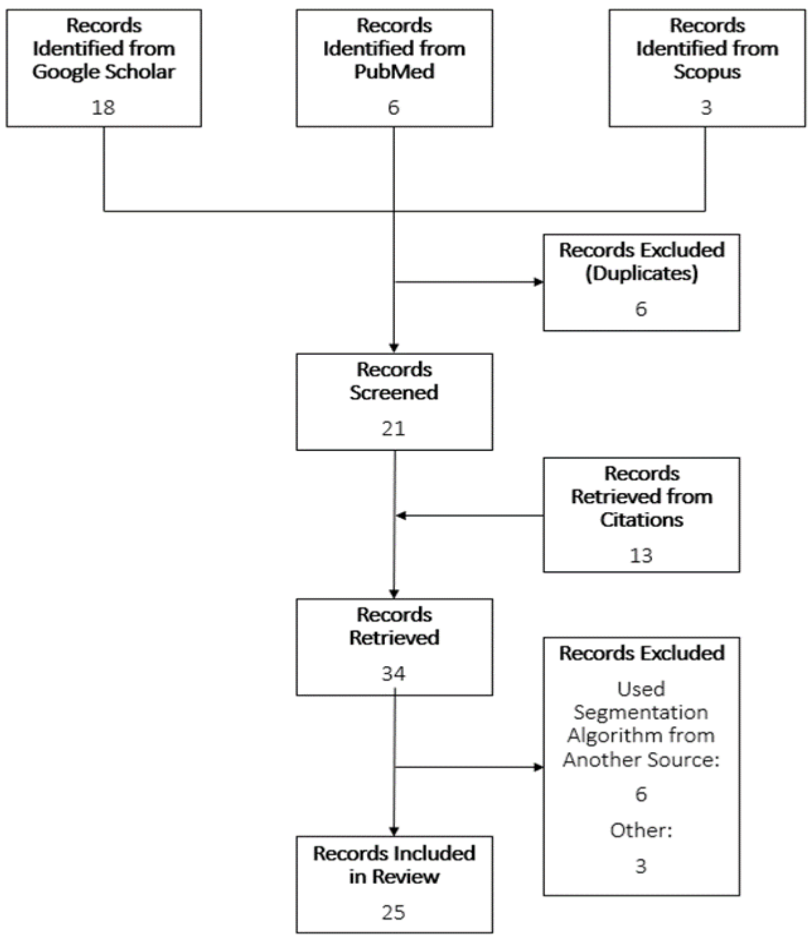


Figure 2.2 Flowchart of Study Selection Process

2.4 Results

Table 2.1 provides a summary of the articles described in this review. Table 2.2 provides a summary of results for each article.

Table 2-1 Summary of articles in this review, sorted by year published.

Source	Test	Population	# Sensors	Sensor Location	Type
Higashi Y, et al. [15]	TUG	10 healthy, 20 hemiplegic	2	Waist, leg	Rule-Based
Weiss A, et al. [23]	TUG	15 healthy adults, 17 PD[1]	1	Lower back	Rule-Based
Salarian A, et al. [24]	TUG	12 healthy , 12 PD[1]	7	Forearms, shanks, thighs, sternum	Rule-Based, Machine Learning
Greene B, et al. [25]	TUG	142 healthy adults; 207 fallers	2	Shins	Rule-Based
Jallon P, et al. [26]	TUG	19 healthy adults	1	Chest	Machine Learning
Adame M, et al. [27]	TUG	10 healthy adults, 20 PD[1]	1	Lower back	Rule-Based
Milosevic, et al. [8]	TUG	4 healthy adults, 3 PD[1]	1	Sternum	Rule-Based
Zakaria N, et al. [28]	TUG	38 elderly	1	Lower back	Rule-Based
Nguyen H, et al. [5]	TUG	16 elderly	17	Head, trunk, hip, scapula, upper arm, forearm, hand, thigh, shin, ankle, foot	Rule-Based
Silva K et al. [18]	TUG	18 elderly	1	Pocket or waist or leg	Rule-Based
Vervoort D, et al. [29]	TUG	59 healthy adults	1	Lower back	Rule-Based
Beyea J, et al. [30]	TUG	10-11 healthy adults	1	Upper back	Rule-Based
Negrini S, et al. [31]	TUG	80 healthy adults (some elderly)	1	Lower back	Rule-Based
Nguyen H, et al. [32]	TUG	12 early stage PD[1]	17	Covering each body segment	Rule-Based
Hellmers S et al. [16]	TUG	148 elderly, 39 healthy young adults	1	Lower back	Machine Learning
Yahalom G, et al. [33]	TUG	25 healthy adults, 25 NPH[2], 15 PD[1]	1	Sternum	Rule-Based
Miller Koop M, et al. [34]	TUG	30 PD[1]	1	Lower back	Rule-Based
De Luca V, et al. [35]	TUG	20 healthy adults	6	Waist, chest, lower back, ankles	Machine Learning
Witchel H, et al. [36]	TUG	23 healthy, 17 MS[3]	2	Lateral lower left thigh, lower right thigh, lower back	Rule-Based

Pew C, et al. [37]	L Test	5 amputees	1	Shank	Machine Learning
Ortega-Bastidas et al. [38]	TUG	25 healthy adults, 12 elderly	1	Lower back	Rule-Based
Hsieh CY, et al. [7]	TUG	5 healthy, 5 OA[4]	4	Waist, Thighs, Wrist	Rule-Based
Hsieh CY, et al. [39]	TUG	26 severe knee OA[4]	6	Chest, lower back, thighs, shanks	Machine Learning
Abdollah V, et al. [19]	TUG	12 healthy adults	1	Head	Rule-Based
Matey-Sanz M, et al. [40]	TUG	5 healthy adults (testing), 1 healthy adult (training)	1	Wrist	Machine Learning

[1]PD = Parkinson's Disease; [2]NPH = normal pressure hydrocephalus; [3]MS = multiple sclerosis; [2]OA = osteoarthritis.

Table 2-2 Results and ground truth comparison methods.

Source	Ground Truth	Stand-up	Sit-down	Turn	Walk	Overall Result
Higashi Y, et al. [15]	Video	r = 0.955	r = 0.735	r = 0.806-0.895	r = 0.809-0.869	
Greene B, et al. [25]	Stopwatch			ρ[1] = 0.83	ρ = 0.89-0.9	
Jallon P, et al. [26]						85% detection
Adame M, et al. [27]	Observational Labelling	Max Abs Err Dev[2] = 0.22 - 0.90 s (healthy) Max Abs Err Dev = 0.19 - 0.81 s (early PD[3]) 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)		
Nguyen H, et al. [5]	Motion Capture	Sn[4] = 100% Sp[5] = 100% Time Diff[6] = 0.03 ± 0.03 s	Sn = 100% Sp = 100% Time Diff = 0.06 ± 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 ± 0.17s	Sn = 100% Sp = 100%

Beyea J, et al. [30]	Motion Capture	RES[7] = -0.01 ± 0.23 to 0.28 ± 0.33s	RES = 0.09 ± 0.24 to 0.38 ± 0.30s	RES = 0 ± 0.30 to 0.27 ± 0.14s	RES = 0.00 ± 0.23 to -0.30 ± 0.21s	Abs Err[8] < ± 0.25s Max Expected Err[9] = 0.34s
Negrini S, et al. [31]	Motion Capture					RMS dev[10] = 0.349 ± 0.196s Avg Bias[11] = 0.183 ±0.212 [-0.372±0.252; 0.738±0.375]
Nguyen H, et al. [32]	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 s to 0.46 ± 0.16 s	Sn = 100% Sp = 100% Time Diff = 0.35 ± 0.16 s
Hellmers S et al. [16]	Stopwatch, Instrumented Chair	Re[12] = 0.84 Pr[13] = 0.66 Acc[14] = 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	
Miller Koop M, et al. [34]	Observational Labelling			r^2 = 0.99		
De Luca V, et al. [35]	Stopwatch	Acc = 87.14		Acc = 66.92%	Acc = 72.3%	Ω_{shA}[15] = 76.47%
Pew C, et al. [37]	Motion Capture			Acc = 96% (SVM) Acc = 93% (kNN) Acc = 91% (ensemble)	Acc = 85% (SVM) Acc = 82% (kNN) Acc = 97% (ensemble)	
Ortega-Bastidas et al. [38]	Video	Avg Err[16] = -0.08± 0.15s (healthy) Avg Err = 0.05 ± 0.30s (older) r = 0.81	Avg Err = -0.01 ± 0.19s (healthy) Avg Err = 0.01 ± 0.29s (older) r = 0.97 - 0.98	Avg Err = -0.08 ± 0.11 to -0.19 ± 0.21 s (healthy) Avg Err = -0.01 ± 0.27s to 0.14 ± 0.62s (older) r = 0.95	Avg Err = 0.17 ± 0.11 to 0.42 ± 0.2s (healthy) Avg Err = 0.19 ± 0.78s to 0.23 ± 0.61s (older) r = 0.89-0.99	
Hsieh CY, et al. [7]	Video	Re = 98 - 100% (healthy)	Re = 94 - 98% (healthy)	Re = 79 - 97% (healthy)	Re = 96 - 99% (healthy)	Re = 94 - 95% (healthy)

		Re = 99% (OA) Pr = 82 - 86% (healthy) Pr = 81 - 83% (OA)	Re = 82 - 86% (OA) Pr = 82 - 88% (healthy) Pr = 94 - 95% (OA)	Re = 92 - 99% (OA) Pr = 96 - 99% (healthy) Pr = 94 - 97% (OA)	Re = 94 - 99% (OA) Pr = 97 - 99% (healthy) Pr = 98 - 100% (OA)	Re = 95 - 95% (OA) Pr = 94 - 95% (healthy) Pr = 94 - 94% (OA) Acc = 95% (healthy) Acc = 95 % (OA)
Hsieh CY, et al. [39]	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. [19]	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. [40]	Video	Mean of Diffs[17] = 0.11s [-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	

[1] ρ = correlation between manual TUG time; [2] Max Abs Err Dev = maximum absolute error deviation; [3] PD = Parkinson's Disease; [4] Sn = sensitivity; [5] Sp = specificity; [6] Time Diff = difference in time; [7] RES = relative error; [8] Abs Err = Absolute Error; [9] Max Expected Err = maximum expected error; [10] RMS dev = root mean squared deviation; [11] Avg Bias = average bias; [12] Re = recall; [13] Pr = precision; [14] Acc = accuracy; [15] Ω_{shA} = percentage of correctly classified participants using the waist accelerometer; [16] Avg Err = average error; [17] Mean of Diffs = mean of differences; [18] RMSE = root mean squared error.

Abdollah et al. [19] 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 [19]. The algorithm published by Beyea et al. [30] was also published in a thesis by Beyea [41].

72% 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 younger adults [16].

Half of the studies that used a single IMU placed the measurement unit on the participant's lower back (eight lower back [16], [23], [27], [28], [29], [31], [34], [38], three chest/sternum [8], [26], [33], four at other locations [18], [19], [30], [40]). Lower back placement approximates whole body center of mass [34]. Beyea [41] originally attempted accelerometer placement on the chest and right shoulder, as well as a goniometer on the knee; however, they later utilized only IMU signals due to the accelerometers being deemed inadequate and the goniometer not providing clearer data than the IMU. Negrini et al. [31] 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 [34], or, static data from each participant used to orient the sensor [2], [16]. Wiess et al. [23] 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 [15], [25], [30], [34], [36]. One study used a cut-off frequency equal to the walking cadence [15]. Some studies utilized low-pass filters with a variety of frequencies [16], [24], [28], [33], [40]. A moving average filter was also used [7], [27], [37], [39]. Multi sensor studies sometimes used a variety of cut-off frequencies for filtering [36], [39]. Nguyen et al. [5], [32] used bandpass filtering. Some studies also chose to normalize data before identifying tasks [30], [31], [32], [32], [38], [40].

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 [24] and another study used a dynamic time warping technique [27]. Both rule-based and machine learning approaches implemented normalization, filtering, and feature calculations during preprocessing, 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 [16], [39]. Calculation of other features from inertial parameters, such as mean, standard deviation, and skewness were sometimes reported [39].

2.4.1 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., at the beginning of the first peak caused by a change in mediolateral angular velocity (ML_{ω})) [8], [27], [30], [38]. A similar processes could be used with the mediolateral rotation (ML_{R}) peak [19], [24], [28], and/or anteroposterior acceleration (AP_{a}) [23], [29], [31]. Beyea et al. [30] used ML_{ω}, AP_{a}, and vertical acceleration (V_{a}) and Silva et al. [18] used ML_{R} and identified consecutive 3° changes to be part of the stand-up task. Another method for labelling 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 [24].

For multi sensor analysis, one study [5] utilized 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 version of the study [32] 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, $ML_{\omega,trunk}$ [7] and $ML_{\omega,waist}$ [15] were used. Some multi sensor studies used an average of each of their ML_{ω} sensors [25], [36].

Most studies identified stand up and sit down tasks using similar methods, and labelled respective tasks based on temporal order [5], [15], [19], [23], [24], [28], [29], [30], [31], [36], [38]. Hsieh et al. [7] identified the sit-down task as when $V_{a,thigh}$ reached a minimum acceleration.

2.4.2 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} [41]. AP_{a} was also used to identify the beginning of walking, with either a threshold [34] or peak identification method [29], [31]. 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 [19], [27], [28], [30], [36], [38]. Nguyen et al. [32] 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.

2.4.3 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 [30], [31], [32], [41] One study that used this method attained a relative error of 0.00 ± 0.30s to 0.27 ± 0.14s [30]. V_{ω} was also used with trunk and head sensors [5] or the waist sensor [7], [15].

Another turn identification method used a threshold value (or observed for increase/decrease [28]) for vertical rotation (V_R) [18], [28], [29], [34], [38], [40]. Also 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 ± -0.21s [38].

A more complex method used V_R with a least squares optimization algorithm to identify turns and reported $r^2 = 0.99$ [34]. Greene et al., identified in their data that $ML_{\omega,shin}$ was lower during turning than walking, and used this relationship to identify turns ($\rho = 0.83$) [25].

2.4.4 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 [16]. As the only hierarchical approach to machine learning subtask segmentation identified in this review, the data was classified into “Static”, “Dynamic”, and “Transition” categories using Boosted Decision trees. Then, the data was further classified into subtasks using Multilayer Perceptions [16]. Jallon et al. [26] 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. [40] proposed a training algorithm using a Multilayer Perceptron model created using TensorFlow Lite, and De Luca et al. [35] proposed an algorithm using k-means clustering. Hsieh et al. [39] 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 [37]. SVM had the highest accuracy for turns at 96%, while Ensemble had the highest accuracy for walking straight at 97% [37].

Seven studies did not compare their algorithm, or individual subtasks, to a ground truth measurement [8], [18], [23], [28], [29], [34], [36]. The results of these studies, therefore, cannot be analyzed.

2.5 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 labelling). 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 between, for example, 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 [7], [16], [19], [35], [37], [39]. 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 [7], [16], [19]. 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. [42], 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 [16], [23], [33]. This is not recommended, since a stopwatch introduces variability and decreases precision in comparison to other methods [41]. If stopwatch data is 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. [16] wanted to work towards 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 is often used as ground truth records in TUG segmentation, providing higher accuracy of measurement than the stopwatch [19], [31], [41]. Beyea [41] used Vicon rather than video recordings due to the ease of processing large quantities of data.

Beyea et al. [30] noted that the calibration and alignment could affect threshold segmentation methods if the data were not first normalized. Salarian et al. [24], [43] also noted that gait and transitions could cause noise, to which a threshold-based algorithm may be sensitive.

2.5.1 Limitations and Criticisms

Numerous studies collected data from less than 15 participants [7], [8], [19], [30], [40], [41]. To avoid type II error, multiple studies indicated the need for an appropriate sample size, which has commonly been cited as at least 20 participants for movement analysis studies [44], [45]. The machine learning algorithm described by Matey-Sanz et al. [40] 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. [33], and there were no further connecting papers related to their process of automatic segmentation. In a study by Koop et al. [34], 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 [28], [31]. In [41], it is not clear which axes are used for vertical and anteroposterior directions.

Instead of a fixed walking distance, Silva et al. [18] 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.

2.6 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 6-minute walk test data [6]. 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.

3 A Smartphone-Based Algorithm for L Test Subtask Segmentation

This chapter addresses the first objective and outlines a novel rule-based algorithm for segmenting the L Test when completed by able-bodied individuals.

The contents of this chapter were published in *BioMedInformatics*:

McCreath Frangakis, A.L.; Lemaire, E.D.; Baddour, N. A Smartphone-Based Algorithm for L Test Subtask Segmentation. BioMedInformatics 2024, 4, 1262-1274. https://doi.org/10.3390/biomedinformatics4020069.

3.1 Abstract

Subtask segmentation can provide useful information from clinical tests, allowing clinicians to better assess a patient's mobility status. A new smartphone-based algorithm was developed to segment the L Test of functional mobility into stand-up, sit-down, and turn sub-tasks. Twenty-one able-bodied participants each completed five L Test trials, with a smartphone attached to their posterior pelvis. The smartphone used a custom-designed application that collected linear acceleration, gyroscope, and magnetometer data, which was then put into a threshold-based algorithm for subtask segmentation. The algorithm produced good results (> 97% accuracy, > 98% specificity, > 74% sensitivity) for all subtasks. These results were a substantial improvement compared with previously published results for the L Test, as well as similar functional mobility tests. This smartphone-based approach is an accessible method for providing useful metrics from the L Test that can lead to better clinical decision-making.

3.2 Introduction

Functional mobility tests are used in rehabilitation to monitor patient progress and assess their ability to move and ambulate safely [3]. Common functional mobility tests include the Timed Up & Go Test (TUG) and L Test of Functional Mobility (L Test) [3]. Although similar in movement, Deathe and Miller [3] reported advantages of the L Test over the TUG, emphasizing that the L Test includes turns in both directions, increases the distance walked, and decreases the ceiling effect often associated with TUG [3].

The L Test begins with the person sitting in a chair. The person then stands up and walks forward three meters to a marker, turns 90°, walks seven meters to a second marker, turns 180°, walks back to the first marker, turns 90°, walks back to the chair, turns 180°, and sits down [3] (Figure 3.1). Each movement is referred to as a subtask (i.e., stand-up, walk, turn, sit-down).

Individual assessment of these subtasks has been useful in predictive recovery measures [5]. For example, fall risk has been correlated with stand-up and sit-down task durations, as well as 180° turns [5]. However, current data collection methods are limited to a clinician's stopwatch to measure total test time [7]; hence, the opportunity to gain extra information is lost. Additionally, clinicians do not have the time to calculate subtask timing from a secondary data collection source within a typical clinical environment. Hence, automating the segmentation and post processing of these tests should enable clinicians to obtain and use subtask information within clinical encounters for decision-making, thereby creating a more effective clinical experience.

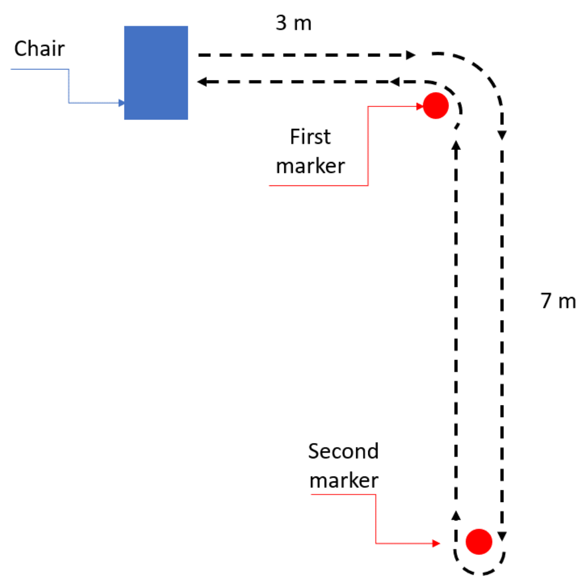


Figure 3.1 Route for the L test. The participant can choose the direction for 180° turns.

A variety of data collection methods have been proposed for subtask segmentation. These include 2D videos recordings, wearable sensors, and ambient sensors [7]. These approaches can also decrease manual error and variance between clinicians when using a stopwatch [7]. Wearable sensors are one of the best options due to their accessibility, lack of space restrictions, and affordability [7]. An inertial measurement unit (IMU) is an inexpensive and accessible type of wearable sensor that can measure parameters such as acceleration, angular velocity, and turn angle. Figure 3.2 shows a physical representation of these parameters, which follow the same labelling convention as [12]: 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). Additionally, the azimuth signal provides the horizontal angle from true north.

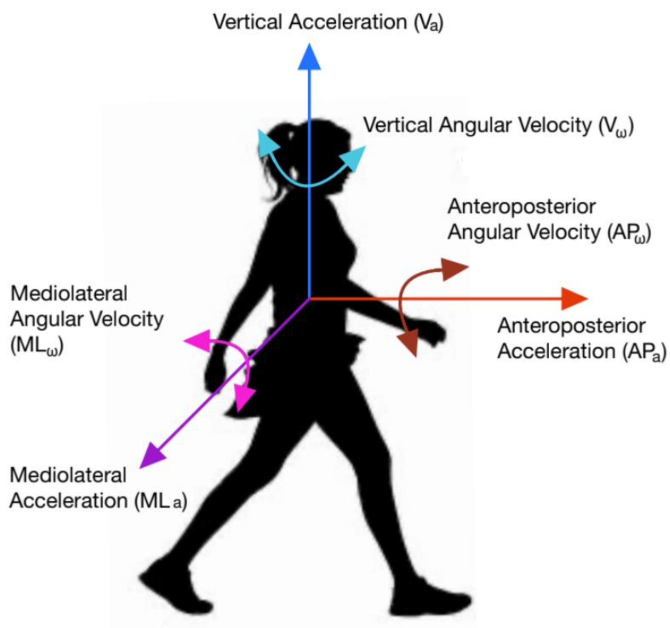


Figure 3.2 Parametric directions used in inertial data.

TUG subtask segmentation using IMU sensors is a valid approach [5], [7], [8], [12], [15], [16], [18], [24], [25], [26], [27], [28], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [46], [47], [48], and can be completed using smartphone sensor data [8], [18], [33]. A smartphone approach does not require additional equipment and has good statistical agreement with movement analysis devices [49]. It was noted that prior smartphone-based studies did not have a primary aim to test the algorithms themselves but rather aimed to establish the validity of using a smartphone IMU for subtask segmentation.

To date, there is no validated approach for segmenting the L Test. In this paper, we report research to develop and evaluate an approach for fully segmenting the stand-up, sit-down, and turn subtasks of the L Test using data acquired from a single pelvis-worn smartphone. Preliminary results were presented in a conference paper [50]. A successful segmentation approach will provide additional movement information for the clinician while requiring equivalent time to complete an L Test trial, accommodating clinical appointment duration. Appropriate subtask segmentation will also enable further analysis and research with AI-based modelling (i.e., fall risk, movement quality, etc.).

3.3 Materials and Methods

3.3.1 Participants

Data were collected from a convenience sample of 6 male and 15 female able-bodied participants, between the ages of 19 and 68 (average age: 36 ± 19 years). Exclusion criteria included individuals with cognitive issues that affected their ability to follow instructions. Participants provided informed consent prior to participating. The study was approved by the University of Ottawa’s Office of Research Ethics and Integrity (H-09-22-8351). Participant characteristics are shown in Table 3.1.

Table 3-1 Participant characteristics

Age Group	Sex	Number of Participants
18-29	Male	3
	Female	9
30-39	Male	0
	Female	1
40-49	Male	0
	Female	0
50-59	Male	2
	Female	4
60-69	Male	1
	Female	1

3.3.2 Data Collection

A custom belt was fastened around the waist of each participant and held an Android smartphone in a posterior pocket (Figure 3.3). This posterior-pelvis position was chosen for its approximation of center of mass, as well as its demonstrated efficacity in other algorithms [51]. A custom app recorded IMU data at 60 Hz; including, raw and linear accelerations in the mediolateral, anteroposterior, and vertical directions; rotation angle in the mediolateral, anteroposterior, and vertical direction; azimuth angle; and angular velocity in the mediolateral, anteroposterior, and vertical directions. Participants were instructed to complete the test at a walking speed that was fast but safe according to their own capabilities. Five trials were completed for each participant. The app provided an auditory cue to indicate that it had begun recording; participants were then informed that they could begin the first trial. Once each trial was completed, the researcher would inform the participant that this was the end of the current trial, and that the participant could begin the next trial. A raw data sample from the app is presented in Figure 3.4.

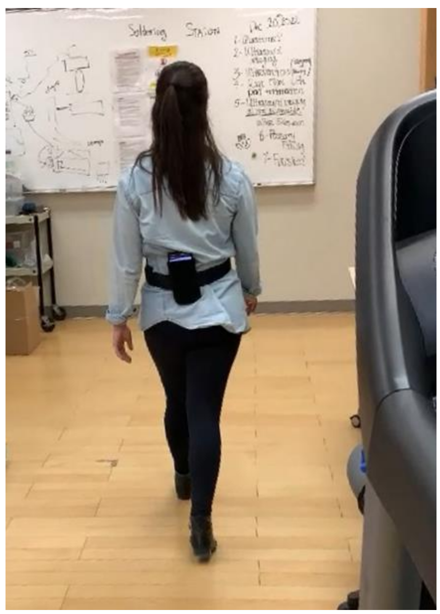


Figure 3.3 Participant completing an L Test trial.

3.3.3 Ground Truth

An Apple iPhone XR was used to video record participants at 30 Hz as they completed the tests. Ground truth time stamps of events of interest were determined from the video using Kinovea [52]. The beginning of the stand-up task was defined as the beginning of trunk flexion and the end corresponded to maximal trunk extension, whether this occurred before or after the first step. Turn initiation was identified as the beginning of pelvis rotation and turn completion was the end of this rotational movement. The beginning of the sit-down subtask was defined as the beginning of trunk flexion, and the end was the end of trunk extension. Timestamps of three foot strikes from the beginning of each trial were also recorded, and then used to align the ground truth times with the inertial data since foot strikes provided clear acceleration peaks.

3.3.4 Preprocessing

Raw data were imported into a custom-built Python program. Initially, the algorithm performed an angle correction on the azimuth signal [51], which corrected jumps in the signal that occur when a participant turns past 360° or 0° [51]. These jumps can be seen in the azimuth plot in Figure 3.4 (c). A threshold technique identified changes in azimuth magnitude greater than 10° between data points and, if true, added or subtracted this magnitude change from the signal [51]. Acceleration and angular velocity data were filtered using a fourth-order zero-lag Butterworth low pass filter with a 4 Hz cut-off frequency [51], a commonly used approach for filtering data for segmentation [12], [30], [51].

3.3.5 Algorithm

The algorithm classified stand-up, sit-down, and turn subtasks. First, the algorithm identified the start of the stand-up task from the mediolateral rotation angle (ML_R) and mediolateral angular velocity (ML_ω) signals. These signals are commonly used in literature for identifying stand-up and sit-down subtasks [5], [7], [8], [16], [18], [24], [25], [26], [27], [28], [30], [31], [32], [35], [36], [38], [39], [40], [47], [48]. A sliding window technique (0.33 s, 20 data points at 60 Hz, overlapping, advancing by one data point) was used to identify the start of these tasks. The average time to complete sitting down or standing up was approximately 1 s, The window size was selected to be sufficiently shorter in duration than the subtask but also large enough to detect relevant changes in magnitude for the target signals. Since the average time to complete the sit down or stand-up task was approximately 1 s, a 0.33 s window was chosen for identifying these tasks.

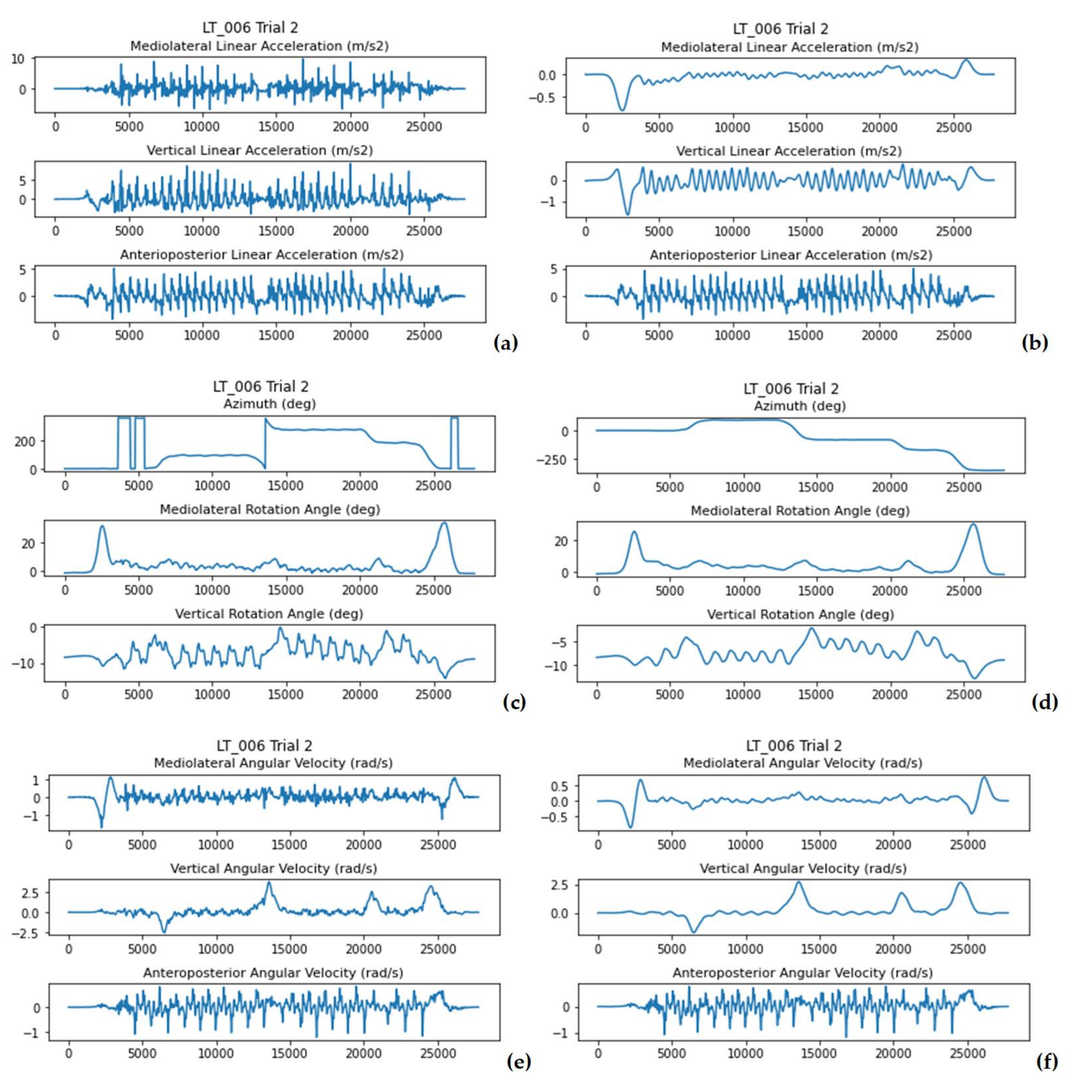


Figure 3.4 Example of raw data collected by the app for (a) linear acceleration, (c) rotation angle, and (e) angular velocity signals; and preprocessed data for (b) linear acceleration, (d) rotation angle, and (f) angular velocity.

The magnitude difference between values at the beginning and end of each window was calculated for both ML_R and ML_ω. The window moved from the start of the data and advanced until a segment was found where both ML_R and ML_ω magnitude differences were greater than their respective threshold values (MDT_{SR}, $MDT_{S\omega}$). Then, the window continued to advance but now calculating the standard deviation (SD) of the values in the window, for both signals. The beginning of standing was the window where the SD for both ML_R and ML_ω crossed their respective thresholds (SDT_{SR}, $SDT_{S\omega}$), set to be the trailing-edge time of the window. Subsequently, the window continued to advance while calculating SD of both ML_R and ML_ω within each window, until the SD for both signals were below new thresholds (SDT_{StR}, $SDT_{St\omega}$). Once this occurred, the leading-edge time of the window was set as the end time for standing up. Thresholds for the stand-up subtask are discussed in section 3.3.6.

Next, the algorithm searched for the first turn subtask. The azimuth signal was used to find turns, and another sliding window approach was used [51]. The window size was changed to two seconds. The window size was determined by finding a sufficient amount of time for participants to complete a full 90° or 180° turn, so that the change in direction could be observed by the azimuth signal, while also being short enough so that the participant had not started the next subtask. Starting from the end of stand-up time, the magnitude difference between the azimuth value at the beginning and end of each window was calculated. Once a window was found where the azimuth magnitude difference was greater than or equal to a threshold (MDT_T), the window would continue advancing, while now calculating the SD of azimuth values within a one second window. The window size used for both standard deviations and magnitude changes was smaller than that in [51], since 90° turns take less time than 180° turns. It was found that a one second window size worked well for the classification of both 90° and 180° turns. The beginning of the first turn subtask was set as the trailing edge time of the window where the azimuth SD was greater than or equal to a threshold (SDT_T). The window then continued to advance while still calculating standard deviations, until the SD was below a threshold (SDT_T). Once this occurred, the leading-edge time of the window where the SD was less than the threshold was set to the end of the first turn subtask. The algorithm then continued to find each turn, using the same process. The sliding window for each subtask would begin at the end time of the subtask that was identified last. Note that the turn before sitting was identified and considered separate from the sitting subtask.

To identify the sit-down subtask, the algorithm used the same process as the stand-up subtask; however, the sliding window started from the end of the data for the given trial and moved 'backwards' towards the second 90° turn. During the L Test, the last 180° turn and the sit-down task are commonly completely simultaneously. Therefore, to ensure that the beginning of the sit-down subtask was within the array used, the sliding window started from the end of the array (i.e., the latest time) to the beginning of the array (i.e., the earliest time), because the 180° turn occurring at the same time as the sit-down subtask often resulted in the sit-down subtask being identified too early, or not at all. Reversing the direction of the window produced better accuracy and specificity in identifying the beginning of the sit-down task. Thresholds for the sit-down subtask are discussed in section 2.5.1.

3.3.6 Threshold and Signal Selection

The thresholds used in the stand-up and sit-down tasks were based on the mean magnitude difference and SD in ML_R and ML_ω for a sliding window (0.33 s, 20 data points at 60 Hz, overlapping, advancing by one data point) across 4 straight walking strides. The averages for all windows were calculated for each participant and then averaged across all participants. Then, the magnitude difference thresholds for the stand-up and sit-down tasks were set to be three standard deviations from the calculated means. These threshold values identified when magnitude changes in ML_R and ML_ω were greater than the mean magnitude change while walking. For this sample of participants, global average thresholds were determined to be MDT_{SR} = 5° and $MDT_{S\omega}$ = 0.5 rad/s. For the thresholds for standard deviation calculations, five standard deviations from the walking mean was experimentally found to be sufficient to confirm that the start or end of the tasks had begun. For stand-up and sit-down subtasks, thresholds were SDT_{SR1} = 2° and $SDT_{S\omega1}$ = 0.15 rad/s. Additionally, some participants had a slower trunk angular velocity but a greater range, which affected their ability to cross the ML_ω threshold. For these cases, another set of thresholds was introduced with lower values for $SDT_{St\omega2}$ = 0.05 rad/s for stand-up (three standard deviations above the mean) and $SDT_{Si\omega2}$ = 0.1 rad/s for sit-down, with a higher threshold value of SDT_{SR2} = 3° (six standard deviations above the mean) for ML_R for both subtasks, ensuring that movement was still taking place. The difference in ML_ω threshold was found experimentally and was higher for the sit-down task, so as to decrease the effect of movement from the last 180° turn on the beginning of the sit-down subtask. The two sets of thresholds ($SDT_{S\omega1}$ = 0.15 rad/s and SDT_{SR1} = 2°, and $SDT_{St\omega2}$ = 0.05 rad/s or $SDT_{Si\omega2}$ = 0.1 rad/s and SDT_{SR2} = 3°) were used together to identify initiation and completion of stand-up and sit-down subtasks.

Table 3-2 Subtask Identification Parameters.

Subtask	Signal	Beginning of Search	End of Search	Direction of Search	Magnitude Change Threshold	Standard Deviation Threshold
Stand-Up	ML_R [1] ML_ω [2] [5], [7], [8], [16], [18], [24], [25], [26], [27], [28], [30], [31], [32], [35], [36], [38], [39], [40], [47], [48]	Start of data array	End of data array	Start to end of array	3 SD above mean [3]	5 SD above mean [3] or 6 SD above mean (ML_R) and 3 SD above mean (ML_ω)
First 90° Turn	Azimuth [51]	One second after end of stand-up	End of data array	Start to end of array	35°	5°
First 180° Turn	Azimuth	One second after end of first 90° turn	End of data array	Start to end of array	35°	5°
Second 90° Turn	Azimuth	One second after end of first 180° turn	End of data array	Start to end of array	35°	5°
Second 180° Turn	Azimuth	One second after end of second 90° turn	End of data array	Start to end of array	35°	5°
Sit-Down	ML_R [1] ML_ω [2]	One second after end of second 90° turn	End of data array	End to beginning of array	3 SD above mean [3]	5 SD above mean [3] or 6 SD above mean (ML_R) and 3 SD above mean (ML_ω)

[1] Mediolateral Rotation Angle

[2] Mediolateral Angular Velocity

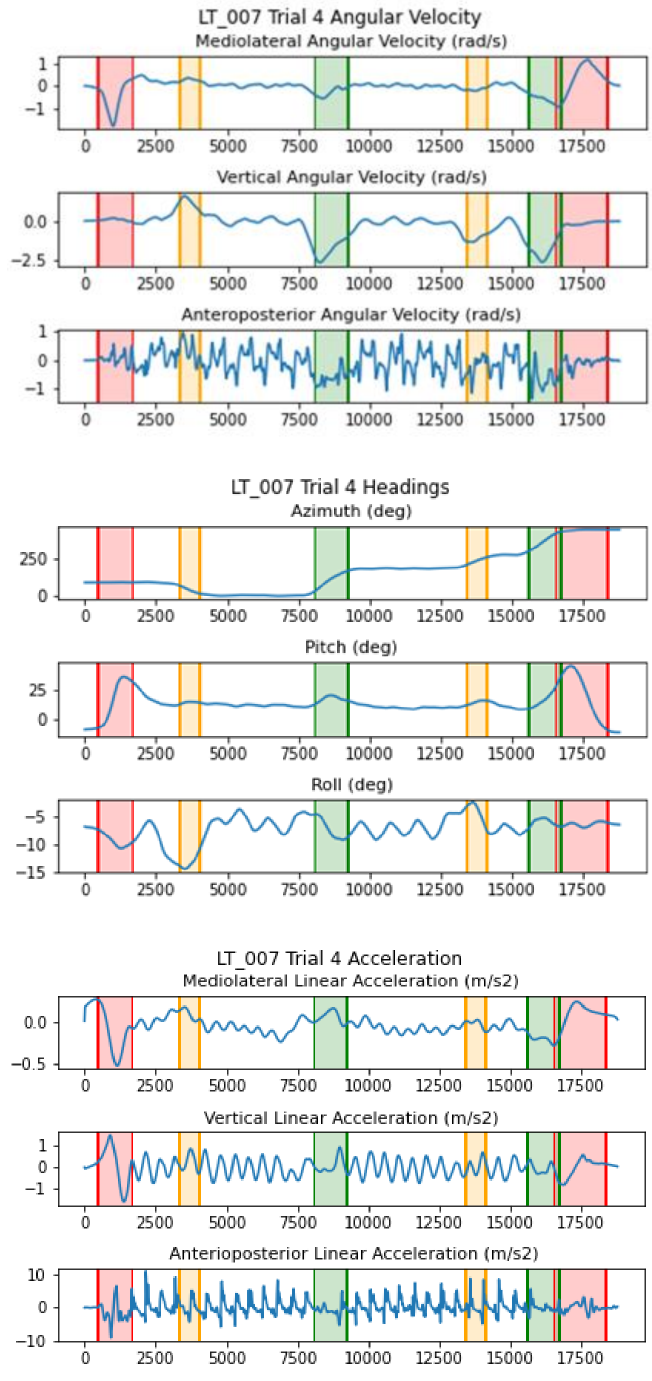
[3] Standard deviations and means are calculated for respective signals for each subtask (section 2.5.2)

The magnitude difference threshold for the turns was set to $MDT_T = 35°$. If this requirement was fulfilled, the SD threshold was set to $SDT_T = 5°$ within a one second window to identify turn initiation and completion. The same thresholds were used to find the end of the turn. The initial thresholds for this task were from [51]; however because the current algorithm must also identify

90° degree turns, the new values were found to be more appropriate for both 90° and 180° turns in the current algorithm. SDT_T, like the stand-up and sit-down subtasks, was approximately five standard deviations higher than when the average person was walking in a straight line. MDT_T was found experimentally using the initial value from [51]. A summary of the identification parameters for each subtask can be found in Table 3.2.

3.4 Results

105 trials from 21 participants were classified by the algorithm. Figure 3.5 shows an example of the output of one trial. Table 3.3 shows ground truth and algorithm-identified means and standard deviations of subtasks for each participant. Table 3.4 shows the accuracy, specificity, sensitivity, and precision for stand-up, sit-down, and turning subtask classification. The performance metrics were calculated using Python's SciKitLearn library, with a 0.06 second (2 video frames) error allowance. Additionally, the table shows the average time differences between the algorithm's estimation of the subtask duration and the ground truth results.



(a)

(b)

(c)

Figure 3.5 Examples of inertial data for an L Test trial with (a) acceleration, (b) rotation angle, and (c) angular velocity signals. Red indicates the stand-up and sit-down subtasks, orange indicates the 90° turn subtasks, and green indicates the 180° turn subtasks.

Table 3-3 Mean and standard deviation durations for all participants.

Participant ID	Stand-Up	GT Stand-Up	Sit-Down	GT Sit-Down	First 90° Turn	GT First 90° Turn	First 180° Turn	GT First 180° Turn	Second 90° Turn	GT Second 90° Turn	Second 180° Turn	GT Second 180° Turn
LT_002	1.47 ± 0.32	0.79 ± 0.44	1.54 ± 0.33	1.05 ± 0.52	0.71 ± 0.42	0.63 ± 0.34	0.89 ± 0.45	1.01 ± 0.51	0.53 ± 0.07	0.61 ± 0.30	0.78 ± 0.18	0.49 ± 0.35
LT_003	1.15 ± 0.12	0.69 ± 0.11	1.6 ± 0.09	1.18 ± 0.19	0.57 ± 0.10	0.69 ± 0.29	0.79 ± 0.07	1.02 ± 0.16	0.52 ± 0.02	0.66 ± 0.17	0.68 ± 0.08	0.72 ± 0.13
LT_004	1.03 ± 0.07	1.07 ± 0.15	1.25 ± 0.08	1.20 ± 0.09	0.60 ± 0.11	0.73 ± 0.08	0.94 ± 0.09	1.10 ± 0.07	0.53 ± 0.03	0.66 ± 0.11	0.82 0.04	1.01 ± 0.06
LT_005	1.35 ± 0.48	0.85 ± 0.32	0.85 ± 0.28	0.74 ± 0.34	0.55 ± 0.02	0.73 ± 0.12	0.88 ± 0.03	1.08 ± 0.08	0.62 ± 0.13	0.87 ± 0.07	0.91 ± 0.13	0.69 ± 0.06
LT_006	1.16 ± 0.10	1.03 ± 0.08	0.57 ± 0.15	1.19 ± 0.10	0.66 ± 0.09	0.95 ± 0.07	1.19 ± 0.11	1.42 ± 0.18	0.59 ± 0.08	0.81 ± 0.14	1.04 ± 0.12	1.02 ± 0.06
LT_007	1.04 ± 0.05	0.89 ± 0.18	1.25 ± 0.29	1.32 ± 0.19	0.60 ± 0.08	0.78 ± 0.16	0.87 ± 0.10	0.98 ± 0.21	0.59 ± 0.08	0.59 ± 0.11	0.77 ± 0.12	0.85 ± 0.12
LT_008	1.02 ± 0.03	1.06 ± 0.12	0.72 ± 0.14	1.43 ± 0.19	0.53 ± 0.04	0.94 ± 0.16	1.28 ± 0.24	1.41 ± 0.13	0.71 ± 0.13	0.88 ± 0.21	0.84 ± 0.17	1.07 ± 0.09
LT_009	1.26 ± 0.15	1.01 ± 0.25	1.5 ± 0.14	1.19 ± 0.14	0.52 ± 0.05	0.70 ± 0.08	0.82 ± 0.10	1.07 ± 0.05	0.5 ± 0.00	0.78 ± 0.16	0.81 ± 0.04	0.96 ± 0.05
LT_010	1.03 ± 0.06	0.82 ± 0.11	1.16 ± 0.43	1.16 ± 0.18	0.66 ± 0.7	0.84 ± 0.09	0.86 ± 0.07	1.15 ± 0.07	0.64 ± 0.07	0.87 ± 0.09	0.83 ± 0.04	1.09 ± 0.09
LT_011	1.00 ± 0.00	0.91 ± 0.07	0.85 ± 0.15	1.14 ± 0.19	0.75 ± 0.12	1.06 ± 0.08	0.94 ± 0.14	1.51 ± 0.20	0.63 ± 0.07	0.89 ± 0.10	0.97 ± 0.12	1.23 ± 0.09
LT_012	1.08 ± 0.09	0.87 ± 0.09	0.96 ± 0.41	1.31 ± 0.20	0.59 ± 0.07	0.62 ± 0.05	0.93 ± 0.10	1.05 ± 0.09	0.78 ± 0.16	0.81 ± 0.10	0.78 ± 0.12	1.01 ± 0.13
LT_013	1.26 ± 0.23	0.96 ± 0.14	1.55 ± 0.38	1.42 ± 0.12	0.61 ± 0.08	0.74 ± 0.12	1.05 ± 0.11	1.11 ± 0.10	0.73 ± 0.11	0.93 ± 0.18	0.78 ± 0.11	0.90 ± 0.08
LT_014	1.25 ± 0.06	0.78 ± 0.06	1.54 ± 0.18	1.14 ± 0.09	0.6 ± 0.12	0.71 ± 0.12	0.99 ± 0.11	1.11 ± 0.05	0.62 ± 0.10	0.82 ± 0.10	0.95 ± 0.15	1.08 ± 0.09
LT_015	1.24 ± 0.14	0.98 ± 0.04	1.53 ± 0.13	1.43 ± 0.24	0.70 ± 0.04	0.86 ± 0.17	1.05 ± 0.18	1.44 ± 0.04	0.72 ± 0.03	1.12 ± 0.04	0.99 ± 0.03	1.30 ± 0.03
LT_016	1.65 ± 0.42	1.55 ± 0.36	1.48 ± 0.36	1.76 ± 0.24	0.84 ± 0.15	0.93 ± 0.10	1.15 ± 0.21	1.39 ± 0.26	0.90 ± 0.19	0.99 ± 0.12	1.16 ± 0.21	1.47 ± 0.27

LT_017	1.31 ± 0.16	0.91 ± 0.12	1.71 ± 0.08	1.43 ± 0.08	0.74 ± 0.06	0.87 ± 0.12	1.03 ± 0.10	1.26 ± 0.04	0.74 ± 0.06	0.80 ± 0.17	1.03 ± 0.10	1.17 ± 0.13
LT_018	1.12 ± 0.10	0.74 ± 0.36	1.44 ± 0.29	1.07 ± 0.53	0.63 ± 0.10	0.77 ± 0.38	0.71 ± 0.36	0.97 ± 0.48	0.68 ± 0.04	0.88 ± 0.43	0.72 ± 0.35	1.06 ± 0.52
LT_019	1.13 ± 0.05	0.96 ± 0.07	1.48 ± 0.33	1.45 ± 0.05	0.60 ± 0.06	0.80 ± 0.07	1.40 ± 0.10	1.50 ± 0.13	0.71 ± 0.09	0.97 ± 0.11	0.98 ± 0.06	1.29 ± 0.10
LT_020	1.04 ± 0.06	0.88 ± 0.05	1.50 ± 0.20	1.25 ± 0.16	0.50 ± 0.00	0.66 ± 0.12	1.14 ± 0.39	1.47 ± 0.23	0.56 ± 0.08	0.67 ± 0.05	0.65 ± 0.03	0.80 ± 0.07
LT_021	1.05 ± 0.05	0.82 ± 0.08	0.83 ± 0.23	1.08 ± 0.07	0.64 ± 0.12	0.95 ± 0.09	0.95 ± 0.11	1.31 ± 0.12	0.66 ± 0.13	0.88 ± 0.15	0.85 ± 0.03	0.98 ± 0.07

Table 3-4 Performance metrics for subtasks. Duration difference is mean absolute difference and standard deviation between algorithm time and ground truth time across all participants and trials.

Metric	Stand-Up	Sit-Down	First 90° Turn	First 180° Turn	Second 90° Turn	Second 180° Turn
Accuracy (%)	98.5	97.1	98.7	98.7	98.9	98.8
Specificity (%)	98.6	98.6	99.8	99.9	99.9	99.8
Sensitivity (%)	97.4	78.3	74.3	81.7	77.1	82.1

3.5 Discussion

The new threshold-based algorithm for segmenting the L Test provided excellent outcomes, with all accuracy and specificity results over 97.1%. These results demonstrated that the proposed smartphone-based approach can provide viable subtask segmentation. Sensitivity results were lower for all subtasks, but still demonstrated improvement upon previous results [50].

The sit-down subtask had lower metrics than the other tasks, most likely due to some participants beginning to drop their head and round their shoulders to look down at the chair before and during the turn before sitting, which may have caused the algorithm to occasionally detect some torso flexion and thus misclassify this movement as the start of the sit-down task. As previously mentioned in section 2.4.2., some of this misclassification was addressed using a higher threshold for this part of the algorithm. Additionally, at the end of the turn, participants would often fidget or fall into the chair, causing them to move a bit after they finished the task. For the stand-up task, participants were more commonly completely still, not introducing variable movements into the signal and therefore having a clear distinction between the beginning and end of the task. Reversing the array (as mentioned in section 2.4.2.) minimized the effect of variable movements on sit-down task initiation identification.

The algorithm had lower performance results with participants who took longer to fully extend their torso during the stand-up task. Similarly, those that took longer turns also tended to have lower subtask identification metrics than those with shorter turn duration. Currently, the thresholds used are global thresholds based on all participants. These outliers were therefore not well labelled by the global thresholds. Future studies could investigate participant-specific thresholds, tuned to their linear walking data. This would allow the algorithm to be directly applied to different patient populations while still ensuring sufficient performance.

The current approach improved upon the results from our preliminary study [50]. The results in [50] achieved 97.9% accuracy, 98.5% specificity, and 86.1% sensitivity for stand-up; 94.6% accuracy, 96.2% specificity, and 72.1% sensitivity for sit-down; and 90.2% accuracy, 95.7% sensitivity, and 70.5% specificity for all turns [50]. The most substantial improvements occurred in sensitivity of the stand-up and turning tasks. The stand-up task obtained an increase in sensitivity of 6.2%, with the turns obtaining increases ranging from 3.8% to 11.6% [50].

A study by Abdollah et al. [48] produced good results for stand up and sit down tasks in the single sensor TUG test, however turns were not segmented. With a single tri-axial accelerometer mounted on a participant's head, they obtained 95% accuracy, 100% specificity, and 90% sensitivity during the stand-up task, and 98% accuracy, 100% specificity, and 98% sensitivity with the sit-down task, using a rule-based threshold algorithm [48]. Our results surpassed these accuracy measurements and were within 1.4% of the specificity outcomes. The current algorithm also surpassed the stand-up subtask sensitivity by 7.4%, however fell short by 19.7% for the sit-down subtask. Pew et al. [37] also published algorithms for segmenting some parts of the L Test, which included results for walking and turning subtasks. They used a variety of machine learning algorithms, with the highest accuracy for turning being 96% with SVM [37]. The current algorithm surpasses this for all turns.

Adame et al. [27] published results for a TUG segmentation algorithm that used a Dynamic Time Warping method from a single IMU attached to the participants' back. Healthy participant results showed a maximum absolute error deviation for the stand-up subtask of 0.22 to 0.90 s, a maximum absolute error deviation for the sit-down task of 0.70 to 0.95 s, and a maximum absolute error deviation for turning subtasks of 0.49 to 1.430 s [27]. With these results, the current study also performed better than Adame et al.'s algorithm for all tasks.

A 0.07 second error allowance is lower than the commonly used L Test approach of using a stopwatch, which has been reported to have a measurement error of 0.2 seconds. Yahalom et al. [33] discussed that measurement error using a stopwatch is estimated to be the same for both stop and start times. However, with a stopwatch, this can vary between clinicians. Additionally, a human controlled approach can also be subjective and introduce further variability in components such as when the clinician begins or ends recording (for example, before or after the patient leans back in the chair during the sit-down task), and can also be subject to distractions [33]. Therefore, even with error allowances, a more objective method of measuring these subtasks will provide more consistent measurements.

3.5.1 Limitations and Future Work

One limitation for the study is that the algorithm was only testing on able-bodied individuals. Future work should include validation with people who have mobility deficits since biomechanical signals can differ from able-bodied participants [53]. Additionally, the assembly of a database of segment timings for the L Test should be done, as it can be used by clinicians to interpret outcomes from data segmentation. The implementation of this algorithm into a smartphone app could provide clinicians with this data in an efficient manner.

3.6 Conclusions

A novel method of subtask segmentation was developed and successfully evaluated for the L Test. When compared to published segmentation results for TUG, the performance metrics of the proposed algorithm often surpassed previous outcomes. The smartphone-based approach was chosen due to its accessibility and ease-of-use, so that the outcomes could be seamlessly integrated into a clinical setting. This technology should allow for precise and useful metrics from functional mobility tests such as the L Test and provide a basis for future AI-based outcome measures.

4 Rule-Based Algorithm for L Test Subtask Segmentation Applied to a Lower Limb Amputee Population

This chapter includes a conference abstract that discusses the results of the algorithm presented in the previous chapter, applied to a lower limb amputee population. This research was presented at the 2024 Congress for the European Society of Physical and Rehabilitation Medicine (ESPRM) and contains results important to the research which contributed to further development of algorithms.

McCreath Frangakis A.L.; Baddour N,; Burger H,; Lemaire E.D. A Smartphone-Based Algorithm for L Test Subtask Segmentation Applied to a Lower Limb Amputee Population. Proceedings of the 24th European Congress of Physical and Rehabilitation Medicine (ESPRM), 2024, Ljubljana, Slovenia. 337.

4.1 Background

Subtask segmentation of functional mobility tests can provide useful outcome measures such as fall risk and movement quality in patient populations. A novel algorithm to segment the L Test of Functional Mobility into stand-up, sit-down, walk, and turn subtasks has previously been designed and showed good results on an able-bodied population [50]. The algorithm uses a sliding window technique to move through the data and identify where the magnitude change and standard deviation have exceeded a threshold set to be five standard deviations above the mean of applicable signals across four walking strides. The original algorithm produced satisfactory results (> 97% accuracy, > 98% specificity, > 74% sensitivity, > 79% precision, and < 355 ± 0.237 s duration error) for all subtasks; however, the algorithm must be assessed for clinical relevance before using this method on data from a lower-limb amputee population.

4.2 Aim

The aim of this research was to assess the viability of a previously designed algorithm using data from a lower-limb amputee population.

4.3 Method

Data was collected from a lower-limb amputee population (31 male, 19 female) at the University Rehabilitation Institute in Ljubljana, Slovenia. A smartphone was attached to the participant's posterior pelvis via a custom belt. A custom app on a Samsung Galaxy S10+

Smartphone recorded inertial sensor data at 60 Hz. Each participant completed one trial. Video recordings were taken of the participants completing the L Test and ground truth time stamps of events of interest were determined from the video. Data were synchronized using timestamps of three foot strikes from the video and accelerometer data. The threshold-based model was applied to the pre-processed smartphone data and the subtask transition times were compared to the ground truth.

4.4 Results

The algorithm produced good accuracy (93-97%) and specificity (97-99%) results across all subtasks. However, sensitivity results were much worse than results for the able-bodied test group (33-60%). Qualitative observation of the amputee participant videos showed different strategies for turning and sitting. People with crutches also had a different standing pattern.

4.5 Discussions and Conclusion

This research shows the applicability of an algorithm designed for able-bodied individuals in a lower-limb amputee population. Further research is required to develop new threshold algorithms for the amputee population. Other modelling methods may be required, such as machine learning to accommodate the variety of strategies used to complete each of the subtasks. With the importance of the L Test for amputee outcome measures, this research has potential to enhance clinical decision making.

5 L Test Subtask Segmentation for Lower Limb Amputees Using a Random Forest Algorithm

The previous chapter demonstrated that the rule-based algorithm did not perform well enough when applied to a lower limb amputee population. Hence, a machine learning approach was developed for L Test subtask segmentation for an amputee population. The contents of this chapter have been published in the journal *Sensors*.

McCreath Frangakis, A.L.; Lemaire, E.D; Burger H.; Baddour N. L Test Subtask Segmentation for Lower-Limb Amputees Using a Random Forest Algorithm. Sensors 2024, 24, 15. https://doi.org/10.3390/s24154953

5.1 Abstract

Functional mobility tests, such as the L Test of Functional Mobility, are recommended to provide clinicians with information regarding mobility progress of lower limb amputees. Smartphone inertial sensors have been used to perform subtask segmentation on functional mobility tests, providing further clinically useful measures such as fall risk. However, L test subtask segmentation rule-based algorithms developed for able-bodied individuals have not produced sufficiently acceptable results when tested with lower limb amputee data. A random forest machine learning model was trained to segment subtasks of the L Test for application to lower limb amputees. The model was trained with 105 trials completed by able-bodied participants and 25 trials completed by lower limb amputee participants and tested using a leave-one-out method with lower limb amputees. This algorithm successfully classified subtasks within one foot strike for most lower limb amputee participants. The algorithm produced acceptable results for lower limb amputees to enhance clinician understanding of a person's mobility status (> 85% accuracy, > 75% sensitivity, > 95% specificity).

5.2 Introduction

The L Test of Functional Mobility can provide information regarding a person's ability to safely ambulate and participate in tasks of everyday life, especially for lower limb amputees (LLA) [3]. Within the L Test, patients must complete a circuit where they stand up from a chair, walk to a marker, turn 90°, walk towards a second marker, turn 180°, walk back to the second marker, turn 90°, walk back to the chair, and sit-down (Figure 5.1). Each individual movement (stand-up, sit-down, turn, walk) is considered a subtask. Inertial measurement unit (IMU) instrumented subtask segmentation of similar functional mobility tests has been analyzed in literature [12], with 15 studies assessing wearable-sensor-based subtask segmentation of the Timed Up-And-Go (TUG) [5], [7], [15], [16], [26], [27], [30], [31], [32], [33], [35], [38], [40], [41], [48] and nine studies using subtask segmentation to inform clinical diagnosis or recovery planning [8], [18], [23], [24], [28], [29], [34], [36], [54]. IMU sensors in smartphone devices are ubiquitous, with five of the 24 studies using these integrated sensors [12]. Smartphone sensor data include linear acceleration, gyroscopic acceleration, rotation angle (Figure 5.2), and azimuth (i.e., angle about the vertical axis from true north). Smartphones are a convenient device to use as an IMU because they are easily accessible and can produce relevant results through custom applications such as in [51].

Smartphone positioning on the individual affects movement detection. The most common location for a single IMU is the lower back [16], [18], [23], [27], [28], [29], [31], [34], [38] because this location gives a good approximation of the user's center of mass. Center of mass acceleration from IMU data has good to excellent reliability for tests such as the Sensory Organization Test, and could identify individuals with varying balance conditions [55].

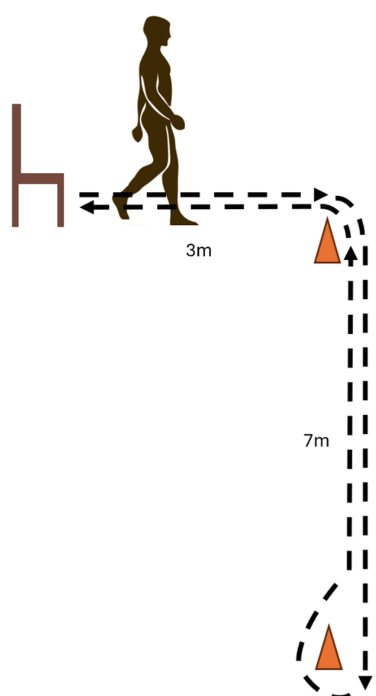


Figure 5.1 Diagram of the L Test.

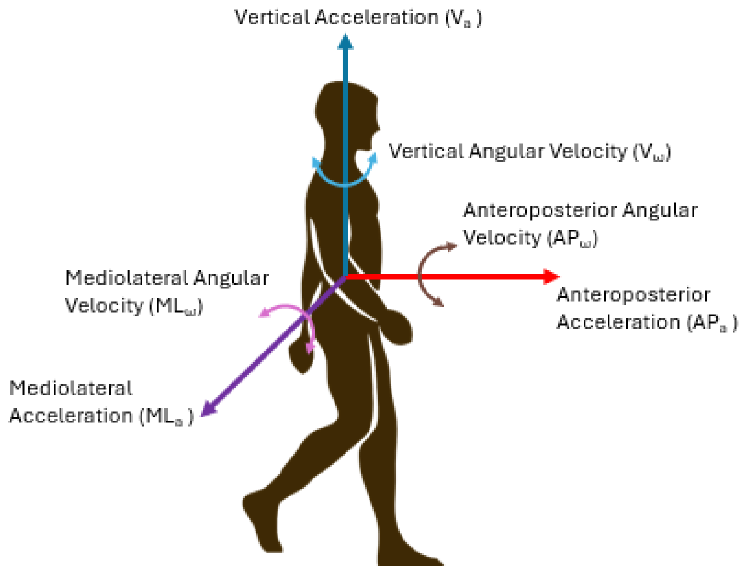


Figure 5.2 Smartphone sensor signal orientation.

While IMU segmentation has been completed for other functional mobility tests, the L Test still requires research on segmentation [12]. Previous research demonstrated good results for rule-based L Test segmentation in able-bodied individuals [50]; however, the algorithm used a generalized approach [56] that did not translate well when tested on a lower limb amputee population [57], which is a recommended population for this test [3]. The rule-based algorithm did not obtain a high sensitivity, likely due to problems classifying the altered, and often slower, mobility pattern that occurs in many lower limb amputees [57]. Therefore, further investigation on L Test subtask segmentation is required for the intended population. More complex algorithms using machine learning approaches may be required to produce suitable results [58]. One other published algorithm partially segmented walk and turn subtasks within the L Test [37]; however, this preliminary research had only five participants who were all unilateral transtibial amputees. Further investigation with a larger dataset is needed to train and assess models segmenting all subtasks for the L Test when completed by lower limb amputees [3].

One study showed good results for fall risk classification from the 6-Minute Walk Test (6MWT) with IMU data and a Random Forest algorithm [6], [59]. This random forest fall risk model worked well with 129 lower limb amputee participants, with a single smartphone IMU. Additionally, Random Forest algorithms have been used to classify wearable IMU data for other applications [60], [61], [62]. These results suggest that a Random Forest algorithm may be a suitable approach for TUG or L Test segmentation for addressing the data from amputee participants.

This paper reports on a novel Random Forest algorithm for segmenting the L Test for people with a lower limb amputation. An algorithm for lower limb amputee L Test analysis could provide a basis for further assessing clinical measures such as fall risk and rehabilitation progress [12]. Information regarding individual subtasks may allow clinicians to identify imbalances during walking or turning and provide an objective outlook on individual ability [12].

5.3 Materials and Methods

5.3.1 Participants

Data were collected from a convenience sample of six male and 15 female able-bodied participants between the ages of 19 and 68 (average age: 36 ± 19 years); and 15 male and 10 female lower limb amputee participants between the ages of 29 and 86 (average age: 58 ± 14 years) with average time since amputation of 11 ± 17 years. Exclusion criteria included individuals with cognitive issues that affected their ability to follow instructions. Participants provided informed consent prior to participating. Able-bodied participant data collection was approved by the University of Ottawa's Office of Research Ethics and Integrity (H-09-22-8351). Lower limb amputee participant data collection was approved by University Rehabilitation Institute of the Republic of Slovenia Soča Ethics Council (035-1/2021-3.4). Participant characteristics are shown in Tables 5.1 and 5.2.

Table 5-1 Able-bodied and amputee participant age and sex characteristics.

Age	Sex	Able-bodied	Amputee
18-29	Male	3	0
	Female	9	1
30-39	Male	0	0
	Female	1	1
40-49	Male	0	0
	Female	0	4
50-59	Male	2	3
	Female	4	3
60-69	Male	1	8
	Female	1	1
70-79	Male	0	1
	Female	0	0
80-89	Male	0	3
	Female	0	0

Table 5-2 Amputee participant characteristics.

Characteristic	Number of Participants
Unilateral	23 (92%)
Bilateral	2 (8%)
Transtibial	16 (64%)
Knee Exarticulation	8 (32%)
Hip Exarticulation	1 (4%)
No Aids	20 (80%)
Single Crutch	5 (20%)

Figure 5.3 Lower limb amputee participant completing an L Test trial.

5.3.2 Data Collection

A custom belt was fastened around the waist of each participant; the belt secured an Android smartphone in a posterior pocket (Figure 5.3). This posterior-pelvis position was chosen for its approximation of center of mass, as well as its demonstrated efficacity in other algorithms [51]. A custom app recorded IMU data at 60 Hz, including raw and linear accelerations in the mediolateral, anteroposterior, and vertical directions; rotation angle in the mediolateral, anteroposterior, and vertical direction; azimuth angle; and angular velocity in the mediolateral, anteroposterior, and vertical directions. Participants were instructed to complete the L Test at a walking speed that was fast but safe according to their own capabilities. Able-bodied participants completed five trials each. Lower limb amputee participants completed one trial each. The app provided an auditory cue to indicate that recording had begun; participants were then informed that they could begin the first trial. Once each trial was completed, the researcher would inform the participant that this was the end of the current trial, and that the participant could begin the next trial.

5.3.3 Ground Truth

A second smartphone (Apple iPhone XR) was used to video record participants at a sample rate of 60 Hz as they completed the tests. Ground truth times of events of interest were determined from the video using VSDC Free Video Editor [63].

Initiation of the stand-up task was defined as the beginning of trunk flexion and termination corresponded to maximal trunk extension, whether this occurred before or after the first step. The reversal of these actions defined the beginning and end of the sit-down task. Turn initiation was identified as the beginning of pelvis rotation and turn completion was the end of this rotational movement.

5.3.4 Data Processing

Raw L Test data from smartphone IMU sensors were imported into a custom-built Python program. Initially, the algorithm performed an angle correction on the azimuth signal [51], which corrected jumps in the signal that occur when a participant turns past 360° or 0° [51]. A threshold technique identified changes in azimuth magnitude greater than 10° between data points and, if true, added or subtracted this magnitude change from the signal [51].

Ground truth timestamps of the initial three foot-strikes of each trial were determined. These were selected due to their clear acceleration peaks, allowing alignment of the ground truth times and inertial data. Acceleration and angular velocity data were filtered using a fourth-order zero-lag Butterworth low pass filter with a 4 Hz cut-off frequency [12], [30], [51].

5.3.5 Datasets

Individual models were trained for each of the six subtests, specifically stand-up, sit-down, first 90° turn, second 90° turn, first 180° turn, and second 180° turn subtasks. To train the six models, the corresponding approximate data ranges were extracted for each subtask as a preprocessing step. Data ranges are shown in Table 5.3. For each subtask, the models are only trained and tested on the corresponding extracted set of data. This subtask data extraction serves to decrease the large class imbalance that occurs when the complete trial data is used in model training. To ensure that important data were not removed, the earliest and latest array index for each subtask were identified across all trials and all participants (Table 5.3). These overall indices were applied to each trial when extracting subtask data. The datasets for each subtask included a ground truth column, which contained one-hot encoded data for the indices for which the subtask was occurring (1) or not occurring (0). After this data extraction, there were six data files for each participant trial: one for each subtask (stand-up, first 90° degree turn, first 180° turn, second 90° turn, second 180° turn, and sit-down). Each of these datasets contained 11 columns: system time; acceleration in the mediolateral, vertical, and anteroposterior axes; angular velocity about the mediolateral, vertical, and anteroposterior axes; rotation angle about the mediolateral and anteroposterior axes; azimuth; and one-hot encoded ground truth.

Table 5-3 Ground truth earliest and latest instance times (by percent of total test time) for all participants (able bodied and LLA).

Subtask	Earliest Instance (%)	Latest Instance (%)
Stand-Up	0	31
First 90° Turn	10	42
First 180° Turn	36	63
Second 90° Turn	61	84
Second 180° Turn	72	97
Sit-Down	76	100

5.3.6 Feature Extraction

Feature sets for mean, standard deviation, variation, maximum value, minimum value, range, kurtosis, and skewness were created for each of the nine IMU signals, using a window size of 24 (0.4 s, overlapping, advancing by one data point) [16], [35], [39]. These features were also calculated for a version of the data set that was reversed, since this was shown to be effective in previous algorithms. Linear acceleration and angular velocity signals were magnitude normalized [12], [40]. The total number of features for each data file was 82. To increase algorithm efficiency, ineffective features were removed using Sci-Kit-Learn's feature selection function, f_classif. This method computes the ANOVA F-value for each feature.

5.3.7 Classification Technique

SciKitLearn's Random Forest Classifier was applied to the datasets for each subtask [64]. Each subtask had its own classifier applied to the one-hot encoded data, for a total of 6 classifiers. This approach with shortened datasets and individual classifiers for each subtask was chosen to decrease the class imbalance that occurs when the entire data set was used to identify movements. If all data were used to train one subtask, the number of instances that were not the subtask would be much greater, which causes issues for recognizing the minority class. For example, for one participant's stand-up subtask, the number of positive instances was 53 and the number of negative instances was 1198 when training with the entire dataset. By only training on the stand-up portion of the data, the number of negative instances was reduced to 335. Since we always know the subtask order, this can be utilized when defining a series of L test data segments for specific classification (i.e., based on dataset analysis, start and end points for each subtask were appropriately defined that encompassed the subtask, with small overlap between subtasks).

The default maximum depth parameter was used, where all nodes expand until either all leaves are pure, or they contain the minimum number of samples [64]. The default criterion used Gini impurity. A variety of tree sizes (50, 100, 250, 500, 750, 1000) were assessed with each of the feature sets (5, 10, 15, 25, 50, 75, 82) using SciKitLearn's parameter optimization function, GridSearchCV [65]. The GridSearchCV function completes an exhaustive search for the highest accuracy (using sklearn.metric.accuracy_score) after a 5-fold cross validation [65]. The output of this classification method gives the number of trees and corresponding number of features with the highest accuracy. To assess whether a chronologically forward or reversed direction of analysis was better for a particular subtask, the GridSearchCV function was used on both forward and reversed data, and the direction that produced the best results was chosen. To produce accuracy, specificity, and sensitivity outcome measures for each subtask, the algorithm was trained on both lower limb amputee data and able-bodied data but tested using a leave-one-out (LOO) strategy on only the lower limb amputee data (i.e., leave one participant out).

5.3.8 Data Analysis

The criteria for successful classification were to identify the transition between subtasks within one average step time for the participant under testing. The average time step was varied by participant. To determine this period, the average step time for each participant was calculated from the ground truth data. For successful classification, the absolute difference between algorithm and ground truth times must be less than the average step time for that participant. The percentage of participants in which this successful clarification occurred is referred to as Accuracy_{AD}. Additionally, to assess successful classification, overall accuracy, sensitivity, and specificity of each of the models were assessed.

5.4 Results

Table 5.4 shows the $Accuracy_{AD}$ results for the beginning and end of each subtask, along with the output for the optimized number of trees and features from the GridSearchCV function, and the most successful direction of analysis. Table 5.5 shows the ANOVA-F feature correlation results used to assemble the lists of features for algorithm training. Only the top ten features for each subtask are shown, for the direction of analysis mentioned in Table 5.4. Table 5.6 shows the overall accuracy, specificity, and sensitivity for each subtask.

Table 5-4 $Accuracy_{AD}$, depth, trees, feature, and direction of analysis results for each subtask.

Subtask	$Accuracy_{AD}$ (%)	Trees	Number of Features	Direction of Analysis
Start of Stand-Up	96	250	5	Forward
End of Stand-Up	96	250	50	Forward
Start of First 90° Turn	92	1000	10	Forward
End of First 90° Turn	76	1000	10	Forward
Start of First 180° Turn	80	500	25	Forward
End of First 180° Turn	92	500	75	Forward
Start of Second 90° Turn	92	500	5	Reversed
End of Second 90° Turn	72	500	5	Reversed
Start of Second 180° Turn	76	250	5	Reversed
End of Second 180° Turn	84	250	5	Reversed
Start of Sit-Down	92	1000	75	Reversed
End of Sit-Down	92	1000	75	Reversed

Table 5-5 Top 10 features for each subtask using ANOVA-F value.

Stand-Up	First 90 Degree Turn	First 180 Degree Turn	Second 90 Degree Turn	Second 180 Degree Turn	Sit-Down
Linear Acceleration X Minimum	Azimuth Range	Azimuth Range	Azimuth Range	Azimuth Range	Pitch Maximum
Pitch Maximum	Azimuth Standard Deviation	Azimuth Standard Deviation	Azimuth Standard Deviation	Azimuth Standard Deviation	Pitch Mean
Linear Acceleration X	Gyroscope Y Range	Azimuth Variation	Azimuth Variation	Azimuth Variation	Pitch
Pitch Standard Deviation	Gyroscope Y Standard Deviation	Gyroscope Y Maximum	Gyroscope Y Minimum	Gyroscope Y Range	Pitch Range
Pitch Range	Azimuth Variation	Gyroscope Y Range	Azimuth Kurtosis	Gyroscope Y Standard Deviation	Pitch Standard Deviation
Gyroscope X Range	Linear Acceleration Y Standard Deviation	Linear Acceleration Y Standard Deviation	Gyroscope Y	Gyroscope Y Variation	Gyroscope X Range
Gyroscope X Standard Deviation	Linear Acceleration Y Range	Linear Acceleration Y Range	Gyroscope Y Mean	Gyroscope X	Pitch Minimum
Linear Acceleration X Mean	System Time	Gyroscope Y	Gyroscope Y Maximum	Linear Acceleration X Minimum	Gyroscope X Standard Deviation
Linear Acceleration X Range	Gyroscope Y	Gyroscope Y Standard Deviation	Roll	Linear Acceleration X	Pitch Variation
Pitch Mean	Gyroscope Y Maximum	Gyroscope Y Mean	Roll Minimum	Linear Acceleration X Mean	Gyroscope X Variation

Table 5-6 Overall accuracy, sensitivity, and specificity results for all subtasks.

Subtask	Accuracy (%)	Sensitivity (%)	Specificity (%)
Stand-Up	95	92	100
First 90° Turn	97	77	100
First 180° Turn	98	87	98
Second 90° Turn	85	75	95
Second 180° Turn	88	85	99
Sit-Down	98	77	100

5.5 Discussion

A novel algorithm was designed to automatically segment L Test subtasks when completed by individuals with lower limb amputations, using data from smartphone IMU sensors. This algorithm successfully classified subtasks within one foot strike for most participants. The algorithm produced acceptable results to enhance clinician understanding of a person's mobility status.

As with similar subtask segmentation studies, identifying ground truth start and end of subtask transitions was often difficult [5], [27]. For example, the beginning and end of turns, especially for individuals with slower gait, can occur over multiple steps. Therefore, there was ambiguity when defining ground-truth for the beginning and end of tasks. In many studies, including the current study, a single researcher labelled all ground truth time stamps; however, some studies used the average of multiple reviewers. Reliability studies for the TUG typically have standard deviations greater than 2 s, for an average TUG time of 11 s to 14 s, which demonstrates the difficulty in interpreting start and end of activities such as sitting, standing, and turning [66], [67]. For assessing AI classification, being within one participant step is likely sufficient for assessing the person's movements during the subtasks, which gives allowance for this common ambiguity in labelling. For clinical interpretation of these results, future research is needed to evaluate clinically meaningful differences for times greater than one step.

The ANOVA-F feature correlation showed that the most correlated feature for all turn subtasks was the azimuth signal's range, followed by the azimuth signal's standard deviation.

Features derived from the azimuth and gyroscope Y signals were most common for all turning subtasks, while features derived from the linear acceleration x, gyroscope x, and pitch signals were most common for stand-up and sit-down subtasks. This is aligned with parameters used for almost all rule-based or machine learning algorithms for TUG test segmentation [12]. System time only appeared in the top ten features for the first 90° turn. This may have been due to the preprocessing of the data into separate files for each subtask, therefore the time that the tasks were completed was less relevant than if the model was given the full array of data for the entire test, and had to differentiate between, for example, the first and second 90° turns.

This is the first study to segment stand-up, sit-down, and all turn subtasks within the L Test using data from people with lower limb amputations. Previous research that used IMUs for mobility analysis reported differences between able-bodied and amputee populations [68], thereby demonstrating the importance of producing models specifically for this population. The current study showed that the Random Forest algorithm is viable for subtask segmentation of lower limb amputee inertial data. Since no other published research is available for amputee full L Test segmentation the following paragraphs compare outcomes from this study to able-bodied or other population results in the literature in other functional mobility tests.

Pew et al. [37] observed the possibility of segmenting turn intent for the L Test completed by a LLA population using a single IMU. The highest accuracy was from a k-Nearest Neighbour (kNN) model that predicted turns with 76% accuracy. Pew et al. also had 96% accuracy with a support vector machine (SVM) model, however this used an approach described as “Individual training”, where accuracies were obtained from models that were trained on the data from the person under investigation. While this approach can lead to a higher accuracy, it is not realistic in a clinical setting, where patient data would not be available to train a new model for each clinical encounter. The current study was evaluated with a Leave-One-Out method, and obtained higher accuracies than the model assessed by Pew et al.

There were several studies that used a single IMU to segment similar functional mobility tests on varying populations. De Luca et al. [35] trained a k-means clustering algorithm on 2-Minute Walk Test (2MWT) and 5 Times Sit to Stand (5STS) data, with the most successful sensor achieving overall accuracies of 84% for stand-up and 72% for turning (20 healthy adults). Results for sit-down were not discussed. The current study surpassed these accuracy values. Hellmers et al. [16] obtained accuracies of 99% for turn around, sit-to-stand, and stand-to-sit of the TUG using a single IMU instrumented belt. Sensitivity for these subtasks were 78% for turn around, 84% for stand-up, and 94% for sit-down. The current study was able to obtain higher sensitivities for 180° turns, which were the only turns assessed in the Hellmers et al. study. The current algorithm was not able to classify the sit-down task as successfully as in the algorithm designed by Hellmers et al. (148 elderly) [13]. Abdollah et al. [48] used head mounted IMUs with a rule-based approach to segment stand-up and sit-down subtasks and obtained an accuracy of 93% for stand-up and 99% for sit-down; sensitivity of 90% for stand-up and 98% for sit-down; and specificity of 96% for stand-up and 100% for sit-down (12 healthy adults). The current algorithm obtained similar accuracies and specificities, however 21% lower sensitivity.

Overall, the current study was often able to achieve better results than previous studies (with differing populations) for turns. However, our proposed approach was commonly as good as other studies for stand-up and sit-down subtasks.

Several studies demonstrated better results, at the cost of using additional IMUs. This included Nguyen et al. [5] who achieved 100% sensitivity and specificity for all subtasks, using an optimal selection of signals from 17 sensors placed around the body. Hseih et al. [39] investigated a variety of machine learning techniques to classify TUG subtasks. An adaptive boosting algorithm obtained 84.84% sensitivity for the stand-up subtask, 89.07% for the second 180° turn subtask, and 80.84% for the sit-down subtask. A decision tree algorithm obtained 87.42% sensitivity for the first 180° turn. These results were obtained using six IMUs. While the use of multiple IMUs can sometimes provide more sensitive results, it is important to consider the financial, accessibility, and setup time requirements of using multiple IMU's. With our proposed approach, all a clinician would need is a smartphone and the time to put a belt on the patient, start the app, and put the app in the belt pocket.

The stand-up and sit-down tasks were the most successful for the Accuracy_{AD} metric, with Accuracy_{AD} scores of 96% for both the beginning and end of the stand-up task and 92% for both the beginning and end of the sit-down subtask. The beginning of the stand-up subtask showed consistent results when analyzed in the forward direction, among all depths, trees, and features. The end of the stand-up subtask gave the best results at a depth of five, and above 250 trees. At higher depths, the algorithm did not perform as well for this subtask.

The first 90° turn was classified with a higher overall Accuracy_{AD} (92% for the beginning of the subtask, 76% at the end) than the second 90° turn (92% for the beginning of the subtask, 72%

at the end). The decrease in accuracy for the end of both turns was most likely due to participants changing their walking angle after the turns. This was noted after the first 90° turn since some participants would walk on a diagonal across the second straight area, to be able to turn their preferred direction for the first 180° turn (e.g., if they start on the right after the turn but want to go around the 180° turn pylon on the left). This diagonal path made selecting the end point of the turn difficult since a distinct angle change did not occur (i.e., was the person still turning or walking along the second straight path). At the end of the second 90° turn, participants may continue to turn slightly to prepare themselves for the 180° and sit-down tasks, therefore decreasing the distinction between the second 90° turn and 180° turn tasks.

The first 180° was classified with a higher $Accuracy_{AD}$ (80% for the beginning of the subtask and 92% for the end of the subtask) than the second 180° turn (84% for both the beginning and end of the subtask). The lower $Accuracy_{AD}$ for the end of the first 180° subtask was likely due to a similar occurrence as for the end of the first 90° subtask, with participants walking diagonally across the straight way to prepare for the subsequent turn. With the second 180° subtask, participants often combined movements for different tasks, resulting in further ambiguity of start and end points. This difficulty in assessing the 180° turn due to surrounding tasks was also seen in other studies, such as [5], where a TUG test with a longer walking time was chosen as the basis for the study's designed algorithm (5 meter instead of 10 meter), since this provided a "gradual transition" between the straight walking and turning, leading to easier evaluation.

5.5.1 Limitations

While this algorithm presents promising insights into the performance of the trained machine learning algorithm, there are a few limitations to consider. The sample size, although larger than most TUG models discussed, should be increased to allow for a more robust model. A larger and more diverse dataset would enhance the reliability and validity of the algorithm, allowing for better representation of the population under study. It would also allow us to look further into how the model may work with new data, as the $Accuracy_{AD}$ performance metric was calculated after hyperparameter and feature optimization. The homogeneity of the sample could also be a limitation. The dataset used in this study comprises primarily similar samples, potentially limiting the algorithm's ability to discern subtle variations or patterns present in more diverse populations. Addressing these limitations could advance the applicability and generalizability of this algorithm and similar algorithms in a clinical setting.

5.6 Conclusions

A Random Forest-based algorithm was developed to segment L Test subtasks, for application to lower limb amputee populations. This novel classifier successfully identified transitions between subtasks within one step, for most lower limb amputee participants. Since all data collection was on a smartphone, this approach could be integrated into a smartphone app that can provide analyses at the point of patient contact. Future research will expand the dataset and explore other deep learning models to improve classification results.

6 Summary, Conclusions and Future Work

6.1 Summary and Conclusions

The objective of this thesis was to create and evaluate algorithms for automatically identifying L Test subtasks for able-bodied individuals and people with lower limb amputations using smartphone sensor data. Initially, a rule-based model was designed to segment the subtask of the L Test using smartphone initial measurement units. The model proposed a general threshold-based approach, which was evaluated with able-bodied individuals and found to give good accuracy and specificity results, with lower sensitivity results. This rule-based model was then evaluated on lower limb amputees, to assess the applicability of the model for different mobility groups. In able-bodied individuals, sensitivity for turns was acceptable and comparable to results from other functional mobility test segmentation algorithms but could be improved with further work. In lower limb amputees, rule-based algorithm sensitivity was low (33-60%), therefore was not able to identify changes in movements as well as many algorithms could with other populations.

To improve outcomes for lower limb amputees, a machine learning (ML) model was adopted, using a random forest algorithm. This ML algorithm was evaluated with lower limb amputees and significantly improved sensitivity results when compared to the rule-based model. With the ML approach, the segmentation results were deemed acceptable for clinical use.

Overall, this thesis contributes to the progression of movement analysis for lower limb amputees, and to the understanding of motions during an L Test.

6.2 Future Work

6.2.1 Walking Segmentation

While the algorithms described in this thesis were able to segment stand-up, sit-down, and turn subtasks, further research could increase the utility of inertial data in a clinical setting. Implementing a step algorithm to assist in assessment, such as in Capela et al.'s 6-Minute Walk Test segmentation [51], would increase the usability of the data. This could allow for more efficient testing that could also provide clinicians with further metrics such as step length and cadence, while contributing to data for fall risk assessment [6].

6.2.2 Data Set

The sample of both able-bodied and lower limb amputee participants included in this thesis resulted in limitations to the generalizability of these results. In the able-bodied participant data collection, 71% of participants were female and 57% were in the under-29 age range. While the sample size was still larger than other similar studies [12], participants of varying age ranges, especially in the 40–70-year-old range, would improve the generalizability of the results.

Additionally, fall risk analysis should be further investigated, with a larger dataset of amputees that are identified to have a higher fall risk. While the L Test is recommended for Lower Limb Amputee populations, it may also be beneficial for analyzing movement in individuals with other mobility impairments. Therefore, future work should also be done to gather L Test data in other populations with variable gait, to investigate the performance results and compare the use of the L Test to the TUG in these other populations.

Ensuring that data quality is sufficient for proper analysis is crucial to effectively apply the algorithm. While smartphone sensors have been shown to be valid for functional mobility tests, sensor errors may sometimes arise. This was observed occasionally in the azimuth signal for some participants, and lead to segmentation difficulty. Therefore, additional error checking methods should be implemented to ensure that the data is of the proper quality to be interpreted by the algorithm. This would decrease the probability of inaccurate results and would allow for clinicians to immediately redo the test if necessary.

6.2.3 Clinical Implementation

Before the proposed algorithm can be used in a clinical setting, it must be transferred to a platform that is both accessible and intuitive for clinicians to use. Smartphone apps such as The Ottawa Hospital Rehabilitation Centre's (TOHRC) Walk Test App already provide a platform with similar functions, such as 2MWT and 6MWT analysis for able-bodied participants [51]. Additionally, further parameters and standards must be investigated and validated for this information to inform clinical procedures. Currently, standard total test times for amputees have been published. However, further data collection including stand-up, sit-down, and turn times for a variety of different amputee populations should be compiled into a data set to gather standards for L Test times, which will inform research decisions. There is further work required to be done to gather fall risk information and provide relative feedback to clinicians regarding patient ability.

References

[1] "Functional Test - EMSCI." Accessed: Dec. 04, 2023. [Online]. Available: https://www.emsci.org/index.php/project/the-assessments/functional-test

[2] D. Podsiadlo and S. Richardson, "The timed 'Up & Go': a test of basic functional mobility for frail elderly persons," *J Am Geriatr Soc*, vol. 39, no. 2, pp. 142–148, Feb. 1991, doi: 10.1111/j.1532-5415.1991.tb01616.x.

[3] A. B. Deathe and W. C. Miller, "The L Test of Functional Mobility: Measurement Properties of a Modified Version of the Timed 'Up & Go' Test Designed for People With Lower-Limb Amputations," *Physical Therapy*, vol. 85, no. 7, pp. 626–635, Jul. 2005, doi: 10.1093/ptj/85.7.626.

[4] F.-Y. Cheng *et al.*, "Factors Influencing Turning and Its Relationship with Falls in Individuals with Parkinson's Disease," *PLOS ONE*, vol. 9, no. 4, p. e93572, Apr. 2014, doi: 10.1371/journal.pone.0093572.

[5] H. P. Nguyen *et al.*, "Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults using multiple inertial sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 36, Apr. 2015, doi: 10.1186/s12984-015-0026-4.

[6] P. Juneau, N. Baddour, H. Burger, A. Bavec, and E. D. Lemaire, "Amputee Fall Risk Classification Using Machine Learning and Smartphone Sensor Data from 2-Minute and 6-Minute Walk Tests," *Sensors*, vol. 22, no. 5, p. 1479, Feb. 2022, doi: 10.3390/s22051749.

[7] C.-Y. Hsieh, H.-Y. Huang, K.-C. Liu, K.-H. Chen, S. J. Hsu, and C.-T. Chan, "Automatic Subtask Segmentation Approach of the Timed Up and Go Test for Mobility Assessment System Using Wearable Sensors," presented at the 2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), New York, NY, USA: IEEE, May 2019. doi: 10.1109/BHI.2019.8834646.

[8] M. Milosevic, E. Jovanov, and A. Milenković, "Quantifying Timed-Up-and-Go test: A smartphone implementation," presented at the 2013 IEEE International Conference on Body Sensor Networks, Aug. 2013. doi: 10.1109/BSN.2013.6575478.

[9] A. Valsangkar, P. Kumar, and E. Scheme, "Automated Segmentation of a Timed Up and Go Test Using an Instrumented Cane," in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Jul. 2021, pp. 1–5. doi: 10.1109/BHI50953.2021.9508595.

[10] P. Savoie, J. A. D. Cameron, M. E. Kaye, and E. J. Scheme, "Automation of the Timed-Up-and-Go Test Using a Conventional Video Camera," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1196–1205, Apr. 2020, doi: 10.1109/JBHI.2019.2934342.

[11] Y. Choi, Y. Bae, B. Cha, and J. Ryu, "Deep Learning-Based Subtask Segmentation of Timed Up-and-Go Test Using RGB-D Cameras," *Sensors*, vol. 22, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/s22176323.

[12] A. L. McCreath Frangakis, E. D. Lemaire, and N. Baddour, "Subtask Segmentation Methods of the Timed Up and Go Test and L Test Using Inertial Measurement Units—A Scoping Review," *Information*, vol. 14, no. 2, Art. no. 2, Feb. 2023, doi: 10.3390/info14020127.

[13] E. Sebastião, B. M. Sandroff, Y. C. Learmonth, and R. W. Motl, "Validity of the Timed Up and Go Test as a Measure of Functional Mobility in Persons With Multiple Sclerosis," *Archives of Physical Medicine and Rehabilitation*, vol. 97, no. 7, pp. 1072–1077, Jul. 2016, doi: 10.1016/j.apmr.2015.12.031.

[14] B. Haas, E. Clarke, L. Elver, E. Gowman, E. Mortimer, and E. Byrd, "The reliability and validity of the L-test in people with Parkinson's disease," *Physiotherapy*, vol. 105, no. 1, pp. 84–89, Mar. 2019, doi: 10.1016/j.physio.2017.11.218.

[15] Y. Higashi, K. Yamakoshi, T. Fujimoto, M. Sekine, and T. Tamura, "Quantitative evaluation of movement using the timed up-and-go test," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 4, pp. 38–46, 2008, doi: 10.1109/MEMB.2008.919494.

[16] S. Hellmers *et al.*, "Towards an Automated Unsupervised Mobility Assessment for Older People Based on Inertial TUG Measurements," *Sensors*, vol. 18, no. 10, p. E3310, Oct. 2018, doi: 10.3390/s18103310.

[17] T. Li *et al.*, "Automatic Timed Up-and-Go Sub-Task Segmentation for Parkinson's Disease Patients Using Video-Based Activity Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 11, pp. 2189–2199, Nov. 2018, doi: 10.1109/TNSRE.2018.2875738.

[18] J. Silva and I. Sousa, "Instrumented timed up and go: Fall risk assessment based on inertial wearable sensors," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2016, pp. 1–6. doi: 10.1109/MeMeA.2016.7533778.

[19] V. Abdollah, T. N. Dief, J. Ralston, C. Ho, and H. Rouhani, "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*, vol. 90, pp. 137–140, Oct. 2021, doi: 10.1016/j.gaitpost.2021.08.020.

[20] G. Sprint, D. J. Cook, and D. L. Weeks, "Toward Automating Clinical Assessments: A Survey of the Timed Up and Go," *IEEE Reviews in Biomedical Engineering*, vol. 8, pp. 64–77, 2015, doi: 10.1109/RBME.2015.2390646.

[21] N. Millor, P. Lecumberri, M. Gomez, A. Martínez-Ramirez, and M. Izquierdo, "Kinematic Parameters to Evaluate Functional Performance of Sit-to-Stand and Stand-to-Sit Transitions Using Motion Sensor Devices: A Systematic Review," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 5, pp. 926–936, Sep. 2014, doi: 10.1109/TNSRE.2014.2331895.

[22] A. C. Tricco *et al.*, “PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation,” *Ann Intern Med*, vol. 169, no. 7, pp. 467–473, Oct. 2018, doi: 10.7326/M18-0850.

[23] A. Weiss *et al.*, “Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson’s disease?,” *Medical Engineering & Physics*, vol. 32, no. 2, pp. 119–125, Mar. 2010, doi: 10.1016/j.medengphy.2009.10.015.

[24] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, “iTUG, a sensitive and reliable measure of mobility,” *IEEE Trans Neural Syst Rehabil Eng*, vol. 18, no. 3, pp. 303–310, Jun. 2010, doi: 10.1109/TNSRE.2010.2047606.

[25] Barry. R. Greene, A. O’Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R. A. Kenny, “Quantitative Falls Risk Assessment Using the Timed Up and Go Test,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2918–2926, Oct. 2010, doi: 10.1109/TBME.2010.2083659.

[26] P. Jallon, B. Dupre, and M. Antonakios, “A graph based method for timed up & go test qualification using inertial sensors,” in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 689–692. doi: 10.1109/ICASSP.2011.5946497.

[27] M. R. Adame *et al.*, “TUG Test Instrumentation for Parkinson’s disease patients using Inertial Sensors and Dynamic Time Warping,” *Biomedical Engineering / Biomedizinische Technik*, vol. 57, no. SI-1-Track-E, pp. 1071–1074, 2012, doi: 10.1515/bmt-2012-4426.

[28] N. A. Zakaria, Y. Kuwae, T. Tamura, K. Minato, and S. Kanaya, “Quantitative analysis of fall risk using TUG test,” *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 18, no. 4, pp. 426–437, Aug. 2013, doi: 10.1080/10255842.2013.805211.

[29] D. Vervoort, N. Vuillerme, N. Kosse, T. Hortobágyi, and C. J. C. Lamoth, “Multivariate Analyses and Classification of Inertial Sensor Data to Identify Aging Effects on the Timed-Up-and-Go Test,” *PLoS One*, vol. 11, no. 6, p. e0155984, 2016, doi: 10.1371/journal.pone.0155984.

[30] J. Beyea, C. A. McGibbon, A. Sexton, J. Noble, and C. O’Connell, “Convergent Validity of a Wearable Sensor System for Measuring Sub-Task Performance during the Timed Up-and-Go Test,” *Sensors*, vol. 17, no. 4, Art. no. 4, Apr. 2017, doi: 10.3390/s17040934.

[31] S. Negrini *et al.*, “Use of Wearable Inertial Sensor in the Assessment of Timed-Up-and-Go Test: Influence of Device Placement on Temporal Variable Estimation,” in *Wireless Mobile Communication and Healthcare*, P. Perego, G. Andreoni, and G. Rizzo, Eds., in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Berlin/Heidelberg, Germany: Springer International Publishing, 2017, pp. 310–317. doi: 10.1007/978-3-319-58877-3_40.

[32] H. Nguyen, K. Lebel, P. Boissy, S. Bogard, E. Goubault, and C. Duval, “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 NeuroEngineering Rehabil*, vol. 14, p. 26, 2017, doi: 10.1186/s12984-017-0241-2.

[33] G. Yahalom *et al.*, “AppTUG-A Smartphone Application of Instrumented ‘Timed Up and Go’ for Neurological Disorders,” Oct. 2018.

[34] M. Miller Koop, S. J. Ozinga, A. B. Rosenfeldt, and J. L. Alberts, “Quantifying turning behavior and gait in Parkinson’s disease using mobile technology,” *IBRO Reports*, vol. 5, pp. 10–16, Dec. 2018, doi: 10.1016/j.ibror.2018.06.002.

[35] V. De Luca, A. Muaremi, O. M. Giggins, L. Walsh, and I. Clay, “Towards fully instrumented and automated assessment of motor function tests,” presented at the 2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI), Mar. 2018, pp. 83–87. doi: 10.1109/BHI.2018.8333375.

[36] H. J. Witchel *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,” *EC Neurol*, vol. 9, no. 10, pp. 689–695, 2018, doi: 10.3389/fneur.2018.00684.

[37] C. Pew and G. K. Klute, “Turn Intent Detection For Control of a Lower Limb Prosthesis,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 4, pp. 789–796, Apr. 2018, doi: 10.1109/TBME.2017.2721300.

[38] P. Ortega-Bastidas, P. Aqueveque, B. Gómez, F. Saavedra, and R. Cano-de-la-Cuerda, “Use of a Single Wireless IMU for the Segmentation and Automatic Analysis of Activities Performed in the 3-m Timed Up & Go Test,” *Sensors*, vol. 19, no. 7, Art. no. 7, 2019, doi: 10.3390/s19071647.

[39] C.-Y. Hsieh, H.-Y. Huang, K.-C. Liu, K.-H. Chen, S. J.-P. Hsu, and C.-T. Chan, “Subtask Segmentation of Timed Up and Go Test for Mobility Assessment of Perioperative Total Knee Arthroplasty,” *Sensors*, vol. 20, no. 21, Art. no. 21, 2020, doi: 10.3390/s20216302.

[40] M. Matey-Sanz, A. González-Pérez, S. Casteleyn, and C. Granell, “Instrumented Timed Up and Go Test Using Inertial Sensors from Consumer Wearable Devices,” in *Artificial Intelligence in Medicine*, M. Michalowski, S. S. R. Abidi, and S. Abidi, Eds., Berlin/Heidelberg, Germany: Springer International Publishing, 2022, pp. 144–154. doi: 10.1007/978-3-031-09342-5_14.

[41] J. B. A. Beyea, “Automating the timed up and go test (tug test) with wearable sensors,” University of New Brunswick, Fredericton, NB, Canada, 2017. [Online]. Available: https://unbscholar.lib.unb.ca/islandora/object/unbscholar%3A8078/

[42] P. Botolfsen, J. L. Helbostad, R. Moe-nilssen, and J. C. Wall, “Reliability and concurrent validity of the Expanded Timed Up-and-Go test in older people with impaired mobility,” *Physiotherapy Research International*, vol. 13, no. 2, pp. 94–106, 2008, doi: 10.1002/pri.394.

[43] A. Salarian, C. Zampieri, F. B. Horak, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian, “Analyzing 180° turns using an inertial system reveals early signs of progress in Parkinson’s Disease,” *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, no. 229, pp. 224–227, 2009, doi: 10.1109/IEMBS.2009.5333970.

[44] A. S. Oliveira and C. I. Pirscoveanu, “Implications of sample size and acquired number of steps to investigate running biomechanics,” *Sci Rep*, vol. 11, no. 1, Art. no. 1, Feb. 2021, doi: 10.1038/s41598-021-82876-z.

[45] S. E. Forrester, “Selecting the number of trials in experimental biomechanics studies,” vol. 2, no. 1, pp. 62–72, Jun. 2015, doi: https://doi.org/10.1080/23335432.2015.1049296.

[46] A. Weiss *et al.*, “Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson’s disease?,” *Medical Engineering & Physics*, vol. 32, no. 2, pp. 119–125, Mar. 2010, doi: 10.1016/j.medengphy.2009.10.015.

[47] D. Vervoort, N. Vuillerme, N. Kosse, T. Hortobágyi, and C. J. C. Lamoth, “Multivariate Analyses and Classification of Inertial Sensor Data to Identify Aging Effects on the Timed-Up-and-Go Test,” *PLoS One*, vol. 11, no. 6, 2016, doi: 10.1371/journal.pone.0155984.

[48] V. Abdollah, T. N. Dief, J. Ralston, C. Ho, and H. Rouhani, “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*, vol. 90, pp. 137–140, Oct. 2021, doi: 10.1016/j.gaitpost.2021.08.020.

[49] S. Mellone, C. Tacconi, and L. Chiari, “Validity of a Smartphone-based instrumented Timed Up and Go,” *Gait & Posture*, vol. 36, no. 1, pp. 163–165, May 2012, doi: 10.1016/j.gaitpost.2012.02.006.

[50] A. L. McCreath Frangakis, E. D. Lemaire, and N. Baddour, “Subtask Segmentation of the L Test Using Smartphone Inertial Measurement Units,” in *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, Jun. 2023, pp. 1–4. doi: 10.1109/BioSMART58455.2023.10162080.

[51] N. A. Capela, E. D. Lemaire, and N. Baddour, “Novel algorithm for a smartphone-based 6-minute walk test application: algorithm, application development, and evaluation,” *J NeuroEngineering Rehabil*, vol. 12, no. 1, p. 19, Feb. 2015, doi: 10.1186/s12984-015-0013-9.

[52] “Kinovea.” Accessed: Sep. 14, 2023. [Online]. Available: https://www.kinovea.org/

[53] N. Capela, E. Lemaire, N. Baddour, M. Rudolf, N. Goljar, and H. Burger, “Evaluation of a smartphone human activity recognition application with able-bodied and stroke participants,” *Journal of NeuroEngineering and Rehabilitation*, vol. 13, Jan. 2016, doi: 10.1186/s12984-016-0114-0.

[54] Barry. R. Greene, A. O'Donovan, R. Romero-Ortuno, L. Cogan, C. N. Scanaill, and R. A. Kenny, "Quantitative Falls Risk Assessment Using the Timed Up and Go Test," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 12, pp. 2918–2926, Dec. 2010, doi: 10.1109/TBME.2010.2083659.

[55] "Use of Mobile Device Accelerometry to Enhance Evaluation of Postural Instability in Parkinson Disease - ScienceDirect." Accessed: Dec. 06, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003999316309856

[56] A. L. McCreath Frangakis, E. D. Lemaire, and N. Baddour, "A Smartphone-Based Algorithm for L Test Subtask Segmentation," *BioMedInformatics*, vol. 4, no. 2, Accessed: May 20, 2024. [Online]. Available: https://www.mdpi.com/2673-7426/4/2/69

[57] A. L. McCreath Frangakis, N. Baddour, H. Burger, and E. D. Lemaire, "A Smartphone-Based Algorithm for L Test Subtask Segmentation Applied to a Lower Limb Amputee Population," in *24th Congress of the European Society for Physical and Rehabilitation Medicine*, Ljubljana, Slovenia.

[58] P. Juneau, E. D. Lemaire, A. Bavec, H. Burger, and N. Baddour, "Automated step detection with 6-minute walk test smartphone sensors signals for fall risk classification in lower limb amputees," *PLOS Digital Health*, vol. 1, no. 8, p. e0000088, Aug. 2022, doi: 10.1371/journal.pdig.0000088.

[59] K. J. F. Daines, N. Baddour, H. Burger, A. Bavec, and E. D. Lemaire, "Fall risk classification for people with lower extremity amputations using random forests and smartphone sensor features from a 6-minute walk test," *PLoS One*, vol. 16, no. 4, p. e0247574, 2021, doi: 10.1371/journal.pone.0247574.

[60] L.-F. Shi, C.-X. Qiu, D.-J. Xin, and G.-X. Liu, "Gait recognition via random forests based on wearable inertial measurement unit," *J Ambient Intell Human Comput*, vol. 11, no. 11, pp. 5329–5340, Nov. 2020, doi: 10.1007/s12652-020-01870-x.

[61] S. Khaksar *et al.*, "Application of Inertial Measurement Units and Machine Learning Classification in Cerebral Palsy: Randomized Controlled Trial," *JMIR Rehabilitation and Assistive Technologies*, vol. 8, no. 4, p. e29769, Oct. 2021, doi: 10.2196/29769.

[62] A. Hua *et al.*, "Evaluation of Machine Learning Models for Classifying Upper Extremity Exercises Using Inertial Measurement Unit-Based Kinematic Data," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2452–2460, Sep. 2020, doi: 10.1109/JBHI.2020.2999902.

[63] "VSDC Free Video Software: Complete Toolset for Video Editing 2023." Accessed: Dec. 06, 2023. [Online]. Available: https://www.videosoftdev.com/

[64] "sklearn.ensemble.RandomForestClassifier," scikit-learn. Accessed: Mar. 25, 2024. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[65] "sklearn.model_selection.GridSearchCV," scikit-learn. Accessed: Apr. 05, 2024. [Online]. Available: https://scikit-learn/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[66] R. C. van Lummel *et al.*, "Intra-Rater, Inter-Rater and Test-Retest Reliability of an Instrumented Timed Up and Go (iTUG) Test in Patients with Parkinson's Disease," *PLOS ONE*, vol. 11, no. 3, p. e0151881, Mar. 2016, doi: 10.1371/journal.pone.0151881.

[67] G. M. Nepal, M. Basaula, and S. Sharma, "Inter-rater reliability of Timed Up and Go test in older adults measured by physiotherapist and caregivers," *European Journal of Physiotherapy*, vol. 22, no. 6, pp. 325–331, Nov. 2020, doi: 10.1080/21679169.2019.1623313.

[68] G. Thibault, E. Lemaire, and N. Baddour, "Amputee Foot Strike Identification Accuracy from 6 Minute Walk Test Raw Data," 2019. doi: 10.13140/RG.2.2.17773.00489.