Automation of the Timed Up and Go Test using an Instrumented Walking Cane

by

Ameya Valsangkar

Bachelor of Engineering, University of Mumbai, 2015

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Science

In the Graduate Academic Unit of Department of Electrical Engineering

Supervisor:	Erik Scheme, PhD, MSc, BSc
Examining Board:	Mahdi S. Hosseini, PhD, Electrical Engineering, Chair
	Julian Meng, PhD, PEng, Electrical Engineering
	Mary E. Kaye, PEng, Electrical Engineering:
	Michael W. Fleming, PhD, Computer Science, UNB

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

October, 2021

Abstract

The Timed Up and Go (TUG) test is used to test a person's mobility and static and dynamic balance. It measures the time a person takes to stand up from a chair, walk three meters, turn around, walk back to the chair, and sit down. Typically, the TUG test is assessed by a physiotherapist with a stopwatch, limiting its effectiveness and making it prone to user error. This has motivated research into automated approaches capable of assessing the various segments of the TUG test using a range of sensing modalities. This study extends upon this body of work by evaluating the feasibility of segmenting the TUG test using an instrumented walking cane. More general contributions are made by introducing the use of error in transition time, as opposed to accuracy, as the cost function during the design of the machine learning framework, and a time-series inspired binary segmentation approach that facilitates the comparison of only two segments at a time. Data was collected using an instrumented cane that measures loading and movement information from 16 participants with musculoskeletal injuries. As a group, the participants yielded TUG times ranging from 11.12s to 28.57s, and a mean of 17.8s. Results of segmenting the TUG test into six segments - Sitting to standing, Walking, Turning, Walking back, Turning back, Standing to sitting - were validated using a leave-one-trial-out and a leave-one-person-out approach, to test both within- and across-participant performance. Various approaches were explored, including conventional classifiers Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM), and extended time series and deep learning methods such as Hidden Markov Models (HMM), CNN LSTM (CLSTM) and Encoder-Decoder Temporal Convolutional Networks (EDTCN). A binary segmentation approach leveraging the temporal nature of the TUG test was adopted with a Dynamic Time Warping (DTW)-based postprocessing alignment. The calculated segmentation error for every case was recorded as both the performance measurement and the optimization parameter as opposed to the traditional use of accuracy of prediction. The results promisingly suggest that the segments or subtasks of a TUG test can be extracted using data collected from a smart cane, laying the groundwork for its automation.

Dedication

This thesis is dedicated to my mother, my family and friends and all my teachers who helped me get to my masters degree.

Acknowledgements

I am extremely grateful to my supervisor Dr. Erik Scheme for his support and sharing his ways to learn. Thank you for your feedback and encouragement with this work and courses. I would also thank Dr. Pradeep Kumar for his daily feedback and continuously providing me with references and topics to learn while also helping me with any topics that I was trying to learn. I have learned a tremendous amount from both of you during our discussions and from the course instructions. I am also very honored to have worked with some of the best minds through the Institute of Biomedical Engineering (IBME) which is continuously progressing and making its mark on the world for preventive healthcare.

I would also thank the Department of Electrical and Computer Engineering and the rest of the IBME students for sharing their insights throughout the time we have been together. I would like to especially thank Dr. Nitin Seth, Alex Roberts and Rushi Bhatt for helping me with the data collection procedure and Dr. Gwyneth DeVries for not only taking out the time to help with the data collection but also for recruiting senior and recovering patient participants, ensuring proper clinical procedure and helping plan out the data collection. Also, a special thanks to Alex Roberts for taking some time out to review my thesis document for language corrections.

A big thanks to my mother and the rest of my family who have always stood by my side even in difficult times and have given utmost value to my education and well being and also to my friends and flatmates Jonas Fernandes and Amit Uppin for

keeping me motivated and hopeful during the difficult period of the pandemic and lockdown.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Gait assessment and clinical gait analysis are essential parts of treating people with balance and mobility issues. They are particularly important for patients recovering from injuries, senior patients with debilitating gait conditions, patients with neurological disorders, and for preoperative assessments [4, 5]. Clinical gait tests improve upon subjective, self-administered assessments by extracting objective information about a person's functional performance and stability during gait. These tests have also been found to be correlated with a person's functional capacity, heart failure and cardiovascular response [6], frailty [7], and fall risk [8]. Falls account for about 40% of all injury-related deaths [9] and are the primary cause of accidental injuries for older adults aged 65 years or more [10]. The fear of falls (or falling again), itself, introduces a vicious cycle where decreased mobility or modified gait may actually increase the risk of a fall [11], further increasing healthcare expenses, decreasing quality of life, and exacerbating balance difficulties.

In response, healthcare professionals are encouraging more proactive gait and mobility screenings to help identify early issues in a person's walking pace, stability, and patterns, to facilitate early intervention and fall prevention. The applications for gait assessments can range from assessing geriatric patients [12] to the rehabilitation of patients after a foot or ankle injury, stroke [13], spinal cord injury [14], or neurological disorder [4, 5]. The Timed Up and Go (TUG) test, which measures the amount of time it takes to stand up, walk 3 meters, and return back to the seat, is a widely accepted medical test for mobility screening [12], frailty assessment [15], and even cognitive skill estimation during gait [16, 17]. When done regularly, the TUG test could aid in proactive prevention programs and early intervention [18]. The standard TUG test typically requires the presence of a clinician who observes the various stages subjectively and times the full test with a stopwatch yielding the overall time as the only recorded objective output. The automation of the TUG test would reduce the burden on the clinician, improve repeatability and generalization of the TUG test, and even alleviate the need to have a clinician present at all. Automation could also facilitate the segmentation of the various phases of the test which may yield added clinical information (e.g. does the subject struggle with the turn, sitting, or sway while walking?) [19]. Furthermore, the automatic detection of the various segments of the TUG test could facilitate their possible extraction from unsupervised data streams recorded during daily living. This could provide added insights about how patients move in their homes and communities, and about how their mobility and balance changes throughout the day and over time. Existing methods, however, must be performed in a controlled or observed environment, are based on heuristics, are expensive to operate, or are inappropriate for adoption by senior populations.

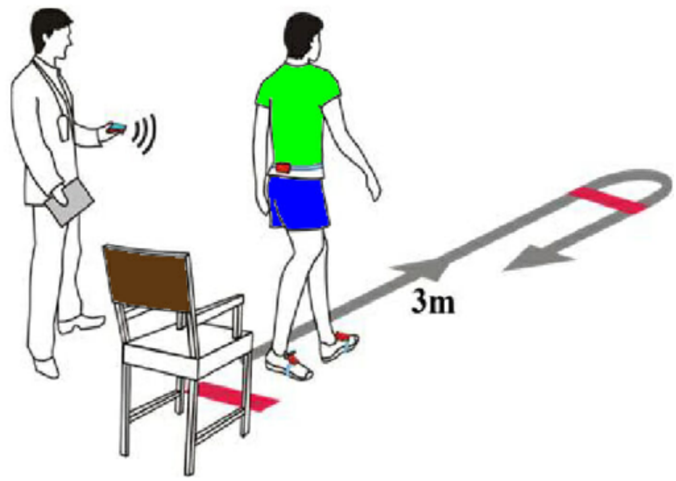


Figure 1.1: Pictorial representation of Timed up and Go (TUG) test and its manual assessment[1].

This work, instead, uses an instrumented smart cane [2], fitted with a 3 axis inertial measurement unit and strain gauges. Instrumenting assistive devices, already used by millions of patients, delivers a natural platform for proactive monitoring and could yield insights about important parameters of gait such as balance, speed, and loading. Such a device is useful for measurements in community dwellings without requiring behavioural change, and eliminates the usual dependence on invasive and lab-intensive testing or equipment. Consequently, this instrumented cane provides a novel mechanism to model and automate the TUG test.

1.2 Background

Mobility and balance in gait

To maintain independence in daily life, mobility is extremely important, especially in senior populations [20]. An estimated 35% of senior adults above the age of 70 suffer from gait disorders. Those with gait disorders are at more than twice the risk of death or institutionalization [21]. Balance and postural stability are known to decline with age but with increased task complexity and the attenuation of sensory feedback, balance impairments can be detected even at relatively earlier ages [20].

Gait analysis tests

Gait tests, generally guided by physiotherapists, are used to assess older adults, patients suffering from injuries, neurological disorders, or rehabilitating after surgery. Many different tests are thus administered to these patients to measure their balance, mobility, fall risk, or recovery. One such test is the 10-meter walk test, where a person walks 10 meters in a straight line and their speed is measured by recording the time. This is in contrast to a 6-minute walk test, where a person walks for 6 minutes and their total distance is recorded. Despite their different approaches and lengths, both can be effective in assessing functional gait in conditions like spinal cord injuries [22]. Similarly, a standing balance test can be useful in measuring postural stability [23]. The TUG test combines both aspects of walking and transferring and is known as a good indicator of quality of gait, fall risk [8] and dynamic postural stability [23].

Need for Automation of TUG test

The TUG test typically requires the presence of a trained clinician to observe the test qualitatively and record the time with a stopwatch. This limits the information that can be extracted from the test and introduces subjectivity in its assessment, leading to poor generalization across days, clinicians, and sites. Conversely, the automation and further segmentation of the TUG test would remove the need for subjective evaluation or even of the presence of a clinician altogether. This could reduce the burden on the healthcare system and increase the number of TUG evaluations that could be completed regularly and proactively outside of a controlled environment. It would also reduce the potential for human error and reliance on specialized training, by making the test more reproducible. In an automated approach, additional information could be gained through the segmentation of the test into its sub-components, namely, 1. Sit, 2. Sit to Stand, 3. Walk (forward and backward), 4. Turn (Before the second walk and before sitting), and 5. Stand to Sit. The performance during each segment could be of clinical importance and may correlate with other tests, thus reducing the number of physical tests required for the assessment of mobility and fall risk. In a fully automated system, the detection of a TUG test could make it possible to extract segments of a TUG test from general activities of daily living, helping clinicians to better understand a patient's gait with regular activity data and a larger number of trials.

Smart cane

This work made use of a smart cane developed in the Health Technologies Lab at the University of New Brunswick (UNB) [2] as seen in Figure 1.2. The use of a smart cane enables non-invasive mobile measurements in community-dwelling environments and makes it easy to track data without the need for an observer. The use of a smart cane also enables the collection of added information about the loading and the balance of the cane user, without requiring them to don wearable sensors. The instrumented cane records and streams strain, accelerometer (x, y, z) and gyroscope (x, y, z) information. These signals can be filtered, windowed, and used as inputs for feature extraction and machine learning.

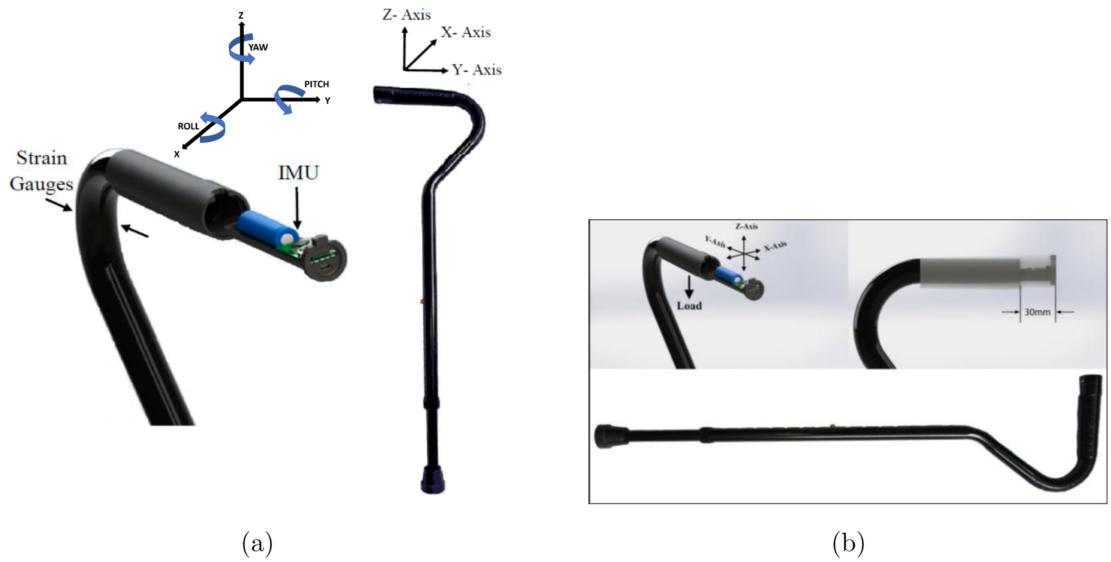


Figure 1.2: Smart assistive cane with embedded IMU sensors and strain gauges [2]

1.3 Objective

This work focuses on the automatic segmentation of a TUG test into its different underlying subtasks. The various phases of the TUG test can be used to assess different functional capabilities which may correlate with other clinical assessments such as the 10 Meter Walk Test, 6 Minute Walk Test or Standing Balance Test. Understanding the relationships between these phases and the impact on gait and motor function they have may facilitate the extraction of added diagnostic information from the TUG test itself [14]. For example, the turning and walking segments require cognitive processing and motor skills and have been linked to the risk of falls [24], [16]. By identifying each segment, future research could explore their correlations with other clinical assessments and the ability to predict TUG metrics from regular gait characteristics. This work aims to automate the assessment of a cane-assisted TUG test to eliminate the need for manual intervention or laboratory environments. This is expected to improve repeatability, save time on testing, provide physicians with valuable patient data without the need for visits, and enable long-term monitoring for proactive healthcare. Automated and pervasive gait monitoring will enable continued assessment of stability, strength, speed, and balance, reducing the need for unnecessary interventions, all while making regular testing more feasible and adaptable to a person's needs. This work also focuses on finding appropriate performance measures for training and optimizing machine learning models. Through this work, it is anticipated that the analytical measurement of gait parameters and quantitative assessment of rehabilitation, fall risk, and mobility will be improved. The major objectives of this work were, therefore, to:

Specific aim 1: Evaluate the feasibility of using a smart instrumented cane to collect and annotate the TUG-test from a target population.

Specific aim 2: Automate the segmentation of the various TUG subtasks, leveraging the known temporal structure of the test.

Specific aim 3: Explore the benefits of using temporal and deep learning architectures to segment the TUG subtasks.

1.4 Thesis Structure

The work in this thesis is organized into six chapters. This first chapter has provided an introduction to the issue being solved, the motivation behind solving the problem, and a preliminary proposed solution. Chapter 2 reviews the justification and feasibility of the proposed work, briefly explaining the requirement for gait analysis, specifically those based on the TUG test, its segmentation, and the use of a smart cane. This further extends to reviewing previous works in the segmentation of the TUG test and the limitations and benefits of these previous approaches. The contents of Chapter 3 explain the approach used in this work to solve the problem, including an extensive data collection procedure in collaboration with a clinician, examples of the collected data, and the developed evaluation framework. Chapter 4 explores the use of conventional classification methods for segmentation, while optimizing the system to improve the results. Chapter 5 provides an exploration of more complex temporal and deep learning techniques that may be applicable given sufficient training data. Chapter 6 concludes the work with a discussion of its major contributions, its limitations, and recommendations for how this work can be built upon in the future.

Chapter 2

Justification and Feasibility

2.1 Gait analysis

Mobility decline remains a significant problem in geriatric populations, patients undergoing rehabilitation, and those with neurodegenerative disorders. Impairment in gait can be an indicator of decreased neurological function in seniors [25] with falls accounting for around 30% of injuries that lead to emergency room admissions in Canada. Falls consequently contribute to $8.7 billion in expenses to Canada's healthcare system each year [26]. Standardized gait and mobility assessments can be used to monitor changes in a person's mobility and balance and to intervene when an increased risk of falls is predicted [27, 12].

Gait itself depends not only on motor function but also on the cognitive ability of an individual [28, 29] and is thus an important predictor for the quality of life and personal independence. An individual's gait pattern may be influenced by age, injuries, and neurological disorders, among others [30]. Consequently, the analysis of gait is important for assessing risk of fall [31], determining balance and mobility [32], measuring quality of life [33], and even in predicting cognitive decline [34]. Walking performance compared to performance on the Stroop test, a classic measure of executive function, and tests of memory show that a healthy gait is associated with higher cognitive function (e.g. catching an object, rather than simple rhythmic tapping). Likewise, dual tasks like 'walking while talking' [35] show slower speeds, longer stride lengths and overall worse gait parameters while reciting given tasks (such as counting backwards), showing again a relation between gait and cognitive function. The epidemiological evidence suggests that, among older adults, fall risk increases in those with impaired neurological conditions. Likewise, the importance of quantification of gait asymmetry and its impact on gait disorders has also been discussed in the literature [36]. This suggests that gait-related issues are highly relevant to an individual's mobility, steadiness in daily tasks and general independence and that proper analysis and segmentation of gait can provide a plethora of information.

Gait segmentation and the underlying information it provides, can thus facilitate the monitoring of mobility, balance, and gait-related progress in recovering patients, older adults with a risk of fall, or patients suffering from disorders such as Parkinson's disease [37]. Gait segmentation may give important information such as stride intervals relating to neurodegenerative diseases [38], stride width to predict risk of falls [39], gait speed relating to Parkinson's disease and for geriatric assessments [40], or stride length for freezing of gait in Parkinson's disease [41], to name a few. Given its clear importance as a diagnostic metric, much information is to be gained from the segmentation and analysis of gait and thus researchers have adopted various sensors and algorithms such as foot switches with heuristic methods [42], inertial sensors and template matching, Dynamic Time Warping (DTW) and hierarchical Hidden Markov Models (HMMs) [37], heuristics on accelerometer and gyroscope signals [43, 44], bipedal foot pressure sensors with HMMs [45], and segmenting of foot-switch signals to identify atypical cycles (cycles that do not match the standard sequence of gait phases) [42] .

In works related to segmentation of gait, the detection of events such as toe-off and heel strike has been explored [42, 46]. Event-based methods have used signal characteristics like peaks, minimums, or zero crossings in the gyroscope or accelerometer signal to identify events [47, 48, 44, 49]. Some researchers have proposed template-based methods to extract individual strides, comparing gait data to a predefined template made from a model stride [50, 51, 52]. Other works include the application of an unsupervised adaptive threshold algorithm to detect changes in signal state [53] and comprehensive gait assessment with derived parameters like stride time, walking speed, and acceleration from inertial sensors to classifying atypical gait cycles [54]. Machine learning methods using k-means clustering, support vector machines (SVM) and neural networks [55], Recurrent Neural Networks (RNN) [56], and Convolutional Neural Networks (CNN) [57] have also been effective in segmenting gait cycles.

Along with gait, a variety of other human activities provide insight into a person's mobility, fall risk and walking ability [58, 59]. Human activity recognition (HAR) has been implemented for various purposes, including surveillance [60], person recognition [61] and clinical gait assessment [62]. HAR has proved important in gait analysis because of its ability to identify the early mobility activities of ICU patients [63] and to evaluate the progress of duchenne muscular dystrophy (DMD) patients [64]. These systems are typically implemented using deep learning methods like CNN on mobile phone data [65], or body sensor and hand gesture data [47], or by using step-wise Linear Discriminant Analysis (LDA) and Hidden Markov Models (HMM) on video based data [66]. Recent studies have developed state-of-the-art human pose estimation techniques to detect key points of the human body based on camera data [67, 68], which are also used in TUG segmentation [69].

As can be seen from the previous literature, gait assessment can be an important tool in the recovery from, or avoidance of, gait impairments. Human gait is highly related to the functions of the nervous system and the musculoskeletal conditions. Gait segmentation allows the detection of abnormal events and assessment of spatio-temporal parameters and their inter-variability during periods of walking [70]. Segmentation of gait provides information related to stride time, speed and sway, while these methods could also be transferable to other clinical gait tests. However, the literature also suggests that gait assessment, balance and mobility analysis should not only entail straight walking, but also turning [71, 72] as well as the transferring aspects of sit to stand and stand to sit [73, 74, 75]. Thus a TUG test, which provides a comprehensive assessment of all these subtasks, may provide rich additional information about a person's functional gait and stability.

2.2 The Timed up and Go Test

Because it combines several sub-tasks in a functionally relevant way, the TUG test has become a preferred clinical assessment for individuals at risk of falls [76]. Statistical testing between functional balance and TUG test completion tasks has found that functional balance may strongly influence the ability to successfully conduct a dual-task TUG in stroke patients [77]. The TUG test also does not suffer from ceiling effects and its results across participants and trials tend to be normally distributed [16]. It has also been found to be reliable in measuring progress in rehabilitation from issues like knee [78] and hip osteoarthritis [79]. The TUG test can be used to detect differences in performance between seniors with and without Parkinson's Disease [4], or to assess the decline of physical mobility with advancing age in otherwise asymptomatic women [80]. Importantly, specific features calculated from the progression of the TUG exhibit significantly different values for fallers and non-fallers [81], suggesting that the automated analysis of TUG measurements could hold substantial diagnostic value. To complete a TUG test, a participant starts by sitting in a chair with no armrests, stands up, walks 3 meters, turns around, walks back and sits back down. In particular, the added turning and transferring aspects of a TUG test can make it more challenging and cognitively demanding than regular ambulation, making it a better indicator for fall risk assessment (87% sensitivity and specificity) than other alternative tests like the Berg Balance Test or Dynamic Gait Index (DGI) [16]. An important observation that gait, turns and turn-to-sit, have shown significant differences between groups with high-fall risks and low fall risks was made in [24]. Specifically, the transitions towards and after a turn in the TUG test are useful indicators of cognition [82] whereas other performance during other segments has been related to frailty [19] and fall risk [81]. The determination of these sub-phases of the TUG test, however, is not feasible in current clinical practice, which relies on stopwatch-based timing of the overall test by a supervising clinician. Consequently, only the total time taken to complete the entire test is typically recorded, missing out on valuable diagnostic information. Automation of such gait assessment tests could reduce dependence on subjective observations and manual timing, and thus improve reproducibility and generalization across patients and centers.

2.3 Use of an instrumented cane

Assistive devices (ADs) are commonly used by people with reduced mobility or balance to provide support and stability and to redistribute loads to compensate for weakened lower limbs [83]. Instrumenting these devices for gait assessment not only alleviates the need for invasive cameras or donning of sensors, but also provides the added ability to measure loading and stability information. Leveraging emerging internet-of-things networks, these devices could also be used to continuously monitor patient progress in the community and at home. Given that nearly a quarter of adults over the age of 65 years use an assistive device and that a 2011 study suggested that 68% of those users use a walking cane [84], a cane is logical choice to explore automating the analysis of gait.

Dahlke et al. [85] suggested that the integration of intelligent assistive technologies, including devices, robotics and sensors in many forms, into eldercare offers opportunities to reduce caregiver burden and enhance healthcare services while improving the quality of life among older adults with mild to severe cognitive deficits. Smart canes are being developed for various purposes, from helping the visually impaired [86], to assessing pain and providing user feedback [87]. An instrumented smart cane, in addition to aiding in balance and confidence, could be beneficial in detecting abnormalities in gait or changes in terrain [88]. A few such devices have been proposed for multi-purpose use that include reminders, feedback about partial weight bearing, pain and progress surveys, and a way to share information about progress between patients and therapists [87]. Others, such as the smartfall [89], were designed for automatic fall detection or to identify gait perturbations for early diagnosis and management of chronic diseases [2]. The ability of these canes to measure timing and speed of placement, angular acceleration, and amounts of weight borne on the cane [90] may prove useful in the remote implementation and analysis of clinical gait testing. Thus, the automatic segmentation of a TUG test using an instrumented cane would make it possible for a physician to monitor and extract a patient's gait characteristics remotely while the patient is in the comfort of their own home.

2.4 TUG segmentation

Because of the untapped potential of automating the TUG test, its segmentation into its various sub-components has been a goal of researchers for many years. This has led to the exploration of a variety of sensors and algorithms, that can largely be categorized into video or sensor-based approaches.

2.4.1 Video-based Segmentation

Video-based TUG segmentation is considered non-invasive because it does not require the placement of sensors on the body, it can be controlled and observed remotely, and the data can be re-watched for added observation. Although it has historically required complex and expensive laboratory equipment such as motion capture cameras, recent advances in deep learning have enabled skeleton extraction from regular video [91]. Still, video-based approaches to TUG segmentation require dedicated camera setups, have a limited field of view and thus a limited operational area, and require a clear line of sight without obstruction from objects like furniture or other people.

Earlier studies made use of the silhouette of the participant from image processing techniques to segment different sub-tasks using specific features like length, width, positions of certain body parts or aspect ratios. For example, in [92], the time required to stand up, and perform the first few gait cycles were calculated in a camera-based study. Likewise, in [93] the local maxima and minima of the head movement was used to detect turns, and the local maxima and minima of the hand position was used to segment the sit to stand and stand to sit subtasks. Such studies had the issue of background noise and required constant view of the involved body part for the algorithms to work. These earlier studies, however, have shown that the assessment of the TUG test can be done remotely by a clinician, given the right surroundings [92, 93, 94].

Over the years, various methods have been developed to heuristically calculate TUG times [95], or locate specific subtasks [96], or all subtasks [97]. To calculate the turn times, a study projected the 3D voxel (digital 3D representation of a person) of a person into a 2D space wherein heuristics were applied to identify the turn [96], but this could depend on observational methods and may not be generalizable. Some authors have tried Kinect-based real-time Range Of Motion (ROM) measurements tracking the X, Y, Z coordinate data to compute the total TUG time [95], but had to modify the TUG test to specifics like having a 10 meter mark and after trying different distances, had to place the camera 4 meters away at a particular angle for the best results. Comprehensive TUG segmentation studies included detecting events of start moving, end rising, start walking, start rotating, start turning, max turn, end turning, end rotating, start lowering, and end moving. Studies that have used skeleton tracking from Kinect sensors [97] have achieved lower errors when detecting these events, but have used few (5) participants, and the segmentation was not fully automated as some of the tracking results had to further be processed manually when segmenting the sub-tasks.

In order to identify more robust and generalizable approaches, studies started exploring machine learning and deep learning strategies. Some exploited the fact that the TUG is largely comprised of simple human poses and made use of human pose estimation followed by Machine Learning (ML) approaches like Long short-term memory (LSTM) and Support Vector Machine (SVM). However, they have found that human pose estimation of the 'Sit' and 'Sit-to-Stand' sub-tasks was difficult due to a lack of sufficient training samples and the 'Turn' subtask was ambiguous in its beginning and ending stages due to very few samples, which led to lower segmentation accuracies [69]. Detecting transition events using HMM [98] on silhouette data had high precision results, but was again conducted in a highly controlled environment with a Kinect camera positioned perpendicularly to the person performing the test, which may not be possible during daily living activities. Another study [91] employed deep learning architectures to extract global 3D poses of the participant and found transition points using heuristic rules. The results reported in this study had lower average segment errors but also acknowledged that the work was camera angle dependent and used only able-bodied subjects. Summarizing, camera systems, although effective, are only practical in labs or dedicated indoor environments and it may be

Table 2.1: Curated video-based works in the automation of the TUG test

#	#Subjects	Device	Methods	Results/Error
[92]	-	camera	active motion pixels	Correlation b/w TUG tasks
[94]	10 healthy	camera	Remote observation	Insignificant diff. in local/Remote TUG
[93]	29 fall hist. 23 without	BioMOBIUS, 2 cameras	max, min in head /hand position	Identified the first turn phase
[96]	7 healthy	2 cameras	Voxel centroid, cosine angle	0.11s error in turn
[97]	5 seniors, 4 young	2 Kinect	max & min	0.1s error
[95]	6 healthy	Kinect	ROM (X,Y,Z)	0.33s TUG error
[69]	24 PD patients	RGB	Human pose estimator, LSTM, SVM, DTW	4.05s TUG error, 0.32-1.07s segment error, 93.1% acc.
[91]	30 healthy	RGB, Kinect	R-CNN, 3D poses	0.15s TUG error
[98]	37 patients	Kinect, 2 stopwatches	HMM	0.001s TUG error

considered inconvenient by some patient populations (particularly geriatric patients) to travel regularly to the controlled environments. A selection of the previous works related to automating the TUG test using video approaches can be seen in Table 2.1.

2.4.2 Sensor-based Segmentation

Unlike cameras, sensors can be used in a variety of spaces, are relatively inexpensive, and provide rich physical information about speed and angular movement during turns. Conversely, wearable devices have so far required careful placement and calibration, precluding their widespread use by many patients.

Many works on sensor-based approaches have used heuristic rules, such as changes in values, or minimum and maximum cutoff value based on the positioning and movement of certain body parts. Early works that combined accelerometer and gyrosensors to identify the activity phases of the TUG test, by applying cutoffs to angular velocities obtained from body worn sensors (like waist worn sensors [99] and thigh worn sensors [100]), found strong correlations with the therapists' observations. This made a strong case for the use of inertial measurement units (IMUs) for the measurement of a TUG test, motivating additional work. Later, a proposed phase segmentation approach based on the minimum and maximum accelerations [101, 102], was found to work well in constrained testing, but may not be generalizable, as the heuristics involved are designed according to the limited available data. Sit-to-stand and stand-to-sit activities, specifically, have further been detected applying specific multi-step custom algorithms to chest mounted gyroscopes [103] and tri-axial accelerometer [104], in different works on balance. Activity defined subtask segmentation has been tightly coupled to the available data and has therefore exhibited limited generalizability due to differences in gait and underlying health problems while performing the said activities [105].

Another popular concept is the use of multi-sensor systems, but this makes the solution very specific and the data collection complicated and ill-suited for daily activities. For example, Nguyen et al. [106] outfitted their participants with 17 inertial motion sensors covering each body part, this is clearly not feasible for everyday use. Nevertheless, in their work, the data were preprocessed to find kinematic peaks for each activity and the time stamps of first minima or maxima to the right or left of these peaks. This methodology, however, may be too specific to generalize across different populations, environments, and conditions of gait. Similarly, another work attempted to segment subcomponents of a TUG test by applying heuristics and their own calculations to a unique subset of sensors to obtain segmentation, and total TUG times [76]. Again, the apparatus involved placing inertial sensors on the forearms, shanks, thighs and sternum, making it difficult to reproduce and likely to violate the conditions under which the heuristics were developed. Ambient sensors have been explored [107], applying cutoffs to sensors including two light barriers, four force sensors, and a laser range scanner built into a single chair apparatus. This offered promising results, but the placement and calibration of the ambient sensors could be demanding and inaccessible to the target populations.

Some have also attempted to use existing mobile devices as a platform; for example, utilizing the built in accelerometers and gyroscopes in an Android-based smart phone, exploiting parameters like changes in angular velocities and zero crossings to identify different subtasks [108]. But this, too, comes with constraints related to the specific device and placement of the device, and required the user to start and stop recording. They were also unable to identify phases of the TUG test automatically in longer sequences [108]. The need for consistent positioning of IMUs remains a challenge in experiments like [109], where the IMU automatically generates a TUG report using dedicated software, which has to be considered as a black box.

An arguably more robust approach of applying machine learning approaches to different types of sensors has also been attempted. Considering the temporal structure of the TUG test, with each segment occurring in sequence, there is more to explore from a time-series perspective. Using IMUs placed on the lower back, [110] used Dynamic Time Warping (DTW) to detect and measure the duration of associated state transitions. Increasing errors were observed, though, with worsening patient condition (healthy, early Parkinson's Disease (PD) and late PD patients), and the positioning of the IMU was critical to detection of the total TUG time. Another work used the dimensionality reduction methods of principal component analysis (PCA) and Independent component analysis (ICA) to reduce the number of features representing useful information from smartphones sensors [111]. While another work segmented TUG phases using HMMs and compared the effect of graph enforcement on different classifiers, the participants were all healthy subjects [112]. The use of IMUs on the lateral side of the shoes coupled, with a two stage classifier, was presented in [113]. The turn was extracted using heuristics and then other subtasks were found by conventional classifiers. This approach, however, failed to leverage the temporal nature of the TUG, required the use of private software for extracting GAIT parameters, and necessitated added procedures like the tapping of the foot 3 times. An overview of some of these previous works can be seen in Table 2.2.

Despite the fairly wide range of research in the automation of the TUG test, the results have been mixed, either with poor performance, or heavily constrained protocols. Consequently, the existing literature motivates the investigation of a different approach in this work.

Table 2.2: Curated sensor based works in the automation of the TUG test

#	#Subjects	Device	Methods	Results
[100]	10 healthy 20 hemiplegic	Gyrocube, Video (GT)	Angular velocities	Correlation, B&A plot ± 1.96 SD
[76]	12 PD, 12 controls	7 waist worn inertial sensors	Analysis of pitch, yaw & angular info.	No significant diff. in total TUG times.
[112]	19 subjects	accelerometer, magnetometer	Roll, pitch, yaw, Graph classifiers	Graph enforced Bayesian (best), detection - 85%.
[107]	5 elderly	Ambient sensors	Cutoffs to 3 types of sensors	Total TUG error 0.05s ± 0.59s.
[101]	23 fallers 18 healthy	Accelerometer (lower back)	min & max accelerations	87% fallers detected
[110]	10 healthy, 10 early PD, 10 adv. PD	IMU in the lower back	DTW based duration assessment	0.06s to 0.4s, 0.06s to 0.38s, 0.14s to 0.32s.
[106]	16 healthy	17 IMUs	Pre-processing, Kinematic peaks	100% TUG transition points
[113]	16 PD	2 shimmer IMUs (shoes)	statistical & signal features 2 stage, turn-heursitics	NB- 56.87%, kNN(k=3)-75.41%, SVM- 81.80%, RF- 75.03%.
[24]	18 seniors	IMU data, Video (GT)	Peak detections to accelerometer & gyroscope signals	Walking & turning most significant between high & low risk fall.
[109]	30 PD	BTS G-Walk (IMU sensor)	Marker positions of the trunk.	No significant diff. in TUG
[105]	5 healthy (h) 5 severe knee osteoarthritis (sko)	OPAL sensors on waist, right & left thigh.	Trunk position relevant to acceleration and angular velocities	Recall, precision, acc h- 94.82%, 94.58%, 95.47%, sko- 94.78%, 94.02% , 95.28%,

Chapter 3

Experiment and Data Analysis

3.1 Data Collection

Experimental data were collected from patients suffering from mobility or musculoskeletal issues, as diagnosed by a collaborating Orthopaedic Surgeon specializing in foot/ankle injuries. Sixteen patient participants were recruited for the study who either already used a standard walking cane, or who were identified as being candidates to use one, either for support due to injury, balance issues, neuro-degenerative diseases, or as a geriatric aid.

The subjects were asked to complete a series of clinical gait tests while using the proposed instrumented cane, including standing balance tests, 10-meter and 6-minute walk tests, and multiple repetitions of the TUG test. Loading and motion data were recorded from the instrumented cane, along with data from a Stepscan© high-resolution pressure sensitive walkway and a set of 6 conventional video cameras. All experiments were conducted under the supervision of a physician, and each subject went through a clinical assessment prior to data collection. Each participant was asked for consent, including the optional collection of video data. All participants consented to be recorded for both the study and for future use of de-identified (face blurred) photos in publications. The experimental protocol was conducted as follows:

- Balance with eyes open - The participant conducted a standing balance test with both eyes open while concentrating on a particular point. This test was conducted twice, once with the instrumented cane and once without the cane. Data were recorded for 10 seconds while standing on a Stepscan pressure tile.

- Balance with eyes closed - The participant conducted the same test as step 1 but with their eyes closed.

- 'Pre' TUG tests- The participant went through multiple TUG tests prior to the more physically demanding walking tests. These tests were conducted using a cane while being recorded by the cameras and walking on the Stepscan tiles.

- 10-Meter Walk tests- Four 10 meter walk tests were conducted using the instrumented cane. The participant walked 10 meters to a marker on the floor, paused for a moment and then walked 10 meters back to the start. This sequence was completed twice.

- 6-minute walk test- The participant was then asked to walk for six minutes while using the cane. The choice of taking either all left or all right turns was left to each participant. In the event that a participant needed to take a break, the time and data stream from the cane were paused. Time and data recording were resumed when the participant started walking again.

- 'Post' TUG tests- The participants performed the same TUG tests as step 3 after the completion of all other tests.

Although the balance test, 10-meter, and 6-minute walk tests were collected for completeness, their analysis was outside of the scope of this thesis, which focuses exclusively on the TUG tests (both pre- and post-). The additional data are being used as part of a larger project in the Health Technologies Laboratory and will be correlated with TUG outcomes in future work. Similarly, it should be noted that the Stepscan pressure data were not used in the analyses of this dissertation, but represent an attractive potential alternative modality for automation of the TUG test in applicable settings.

A summary of the demographic information of the participants is shown in Table 3.1, where it can be seen that most participants were older adults. These participants had gait issues like knee or ankle osteoarthritis, recent lower limb injuries, or some other gait-affecting condition. The affected legs were split equally between being the right and left leg, and the participants were instructed to use the cane as they saw fit, with guidance from the clinician as needed. The participants were equally divided by sex but were primarily right handed. Most participants preferred using the cane in their right hand, irrespective of the affected leg [1] .

Table 3.1: Demographic data distribution across participants

Parameter	Measurement	Value
Gender	Percentage	M-40.90%, F-59.09%
Age	Mean+StdDev	66.55+11.34
Weight (kgs)	Mean+StdDev	84.43+17.67
Height (cms)	Mean+StdDev	186.17+38.97
Affected Leg	Count/ Percentage	42.1% L, 52.63% R, 5.26% Both
Hand Dominance	Count/ Percentage	5.55%L, 94.73%R

3.2 Signal Visualization and Feature Exploration

The cane provides raw strain, accelerometer (x, y, z) and gyroscope (x, y, z) data recorded from the handle of the device. These signals were sampled at a rate of 100Hz and low pass filtered, with a Butterworth filter of 6^{th} order and cutoff frequency of 4 or 8 Hz, to remove high frequency noise. The filtered data includes time, filtered strain,

[1] This is in contrast to best clinical practice, which suggests using the cane in the opposite hand from the affected limb

accelerometer- (x, y, z), gyroscope- (x, y, z). Additionally, pitch, roll, and yaw values were individually derived from angular information, accelerometer, and gyroscope values. The TUG test fundamentally consists of 5 different types of motion and 8 distinct segments, as reflected in Figure 3.1. These segments represent pieces of the complete TUG test in which the subject starts from sitting then stands up, walks, turns, walks back, turns again, and finally sits back down. It can be seen that the strain information in Figure 3.1 denotes the periodicity of loading observed during the walking phase. Depending on how heavily the cane is used to get up from the chair, the first peak observed in the strain, along with the rise and subsequent fall around that peak, may be used to identify the sit-to-stand segment. The acceleration values have higher variance during cane activity, but are more stable during the sitting phases. The gyroscope values show the (angular) turning of the cane in three directions and prove to be important for identifying the turning aspect of the TUG test. Although not shown in Figure 3.1 the angle roll, pitch, and yaw, derived from the raw signals, can also aid in showing the relative rotation of the cane in motion and reaffirm periodicity while walking, spikes when turning, and relative stability. To establish a ground-truth timing of the segments, the cane data were labelled manually, identifying the subtask timings by matching the activities in the corresponding video data. This is consistent with how a physician would observe these gait events in practice in the absence of a recording device, but with the added ability to review and dial in on specific timings for increased accuracy.

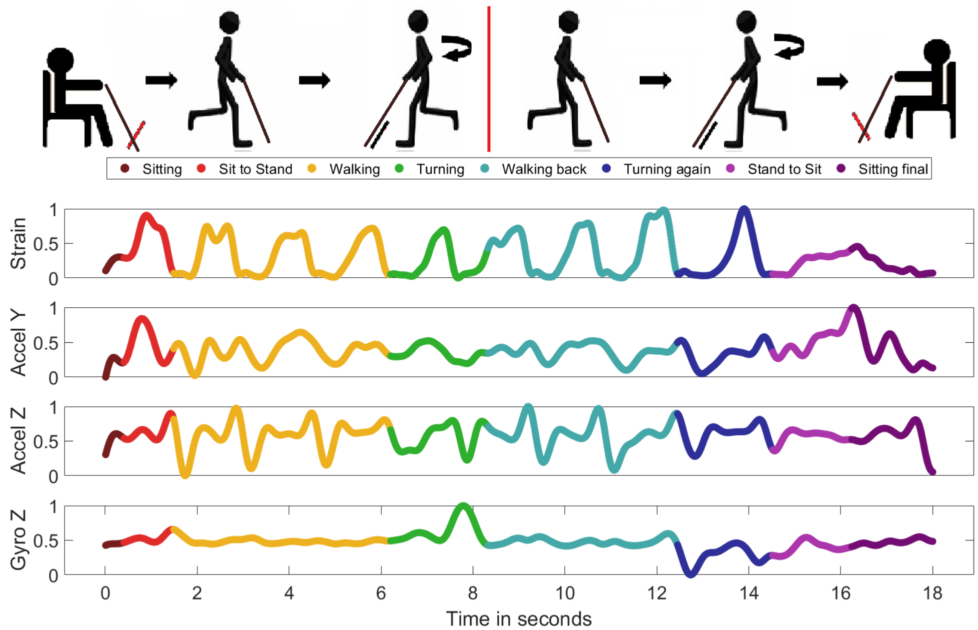


Figure 3.1: Example of the sensor readings during various segments in the TUG test when using the instrumented cane.

Scatter matrix analysis: Initially, the subtasks of sit, walk, and turn, which occur twice each, were labelled with the same identifier, yielding 5 unique subtasks. Subtask labels were thus assigned follows, 0- Sitting (initial and final), 1- Sit to Stand, 2- Walk (ahead and back), 3- Turn, 4- Stand to Sit. However, plotting the signals against each other for these subtasks (as in Figure 3.2) quickly identified differences within them. For example, when the gyroscope-derived yaw is plotted against the acceleration Y-values, and grouped according to these subtasks, it is clear that the two episodes of each subtask are dissimilar. They also indicate that the turn phase, which lasts for a comparatively shorter time, could be separable using a combination of these features. Similarly, for most participants, plotting the gyroscope-derived yaw against the gyroscope Z-values shows clear turns. Strain also shows a potentially separable walk pattern when plotted against the gyroscope yaw values.

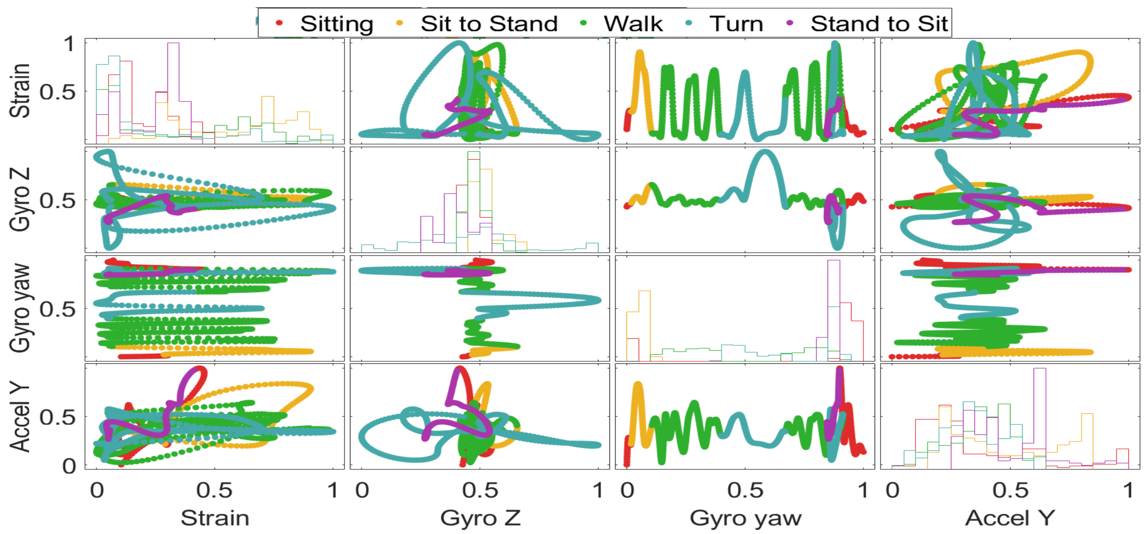
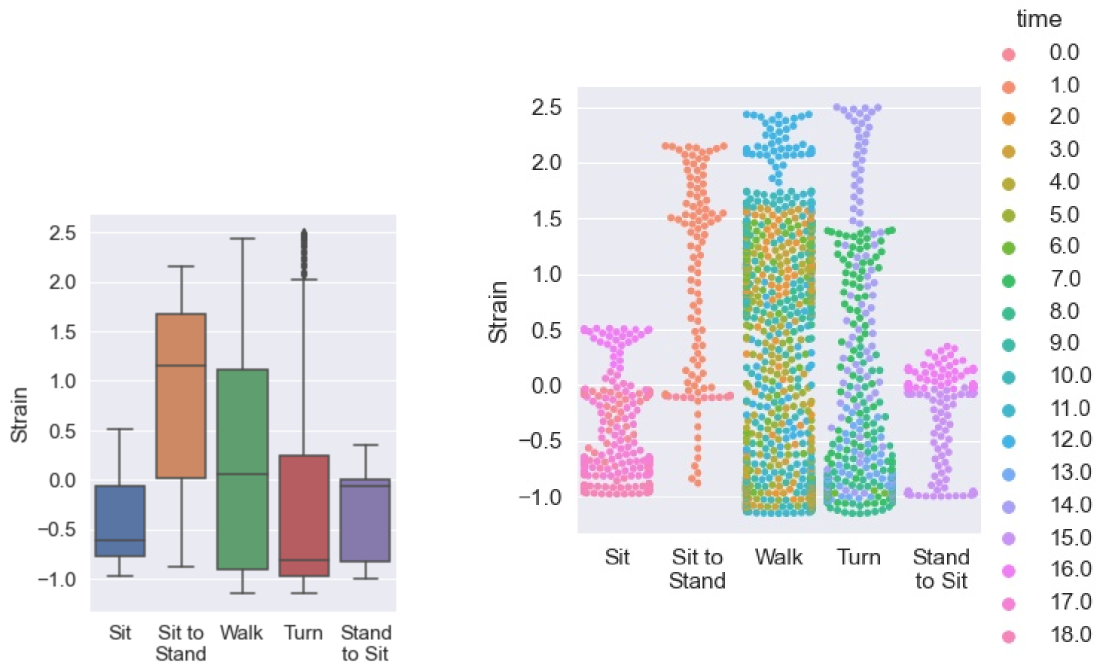


Figure 3.2: Scatter matrix of prominent derived signals

Box- & Swarm plot analysis: Both box and swarm plots were used to explore the averages, data quartiles, skewness, and outliers of the data being analyzed. They were also used to show the distribution of data by time and each subtask. Figure 3.3a shows strain divided by class (subtask), whereas Figure 3.3b shows the strain values colored according to the time interval, in seconds. The boxplot shows that strain values were distributed equally around zero mean, given that the cane is used in a periodic sense for the 'walk' subtask. While 'sitting' and 'turning', the strain values were generally lower than zero because many participants lifted the cane off the ground (thus "pulling" on the handle). Conversely, the sit to stand phase had higher than average strain, as the participants used the cane to stand up. The swarm plot, distributed according to class labels, while keeping time as hue, shows that time plays a major factor on the distribution of strain. In this example, strain is densely concentrated in the early and late times around a normalized 0.0 (sitting, initial and final) and is quite equally distributed around 1 and -1 between time marks of 6-8 seconds and 10-12 seconds (showing the striking and lifting of the cane while walking forward and walking backward, respectively). The second turn starts at the 13$^{\text{th}}$ second with lower strain near 0 and ends with higher strain values around the 13$^{\text{th}}$

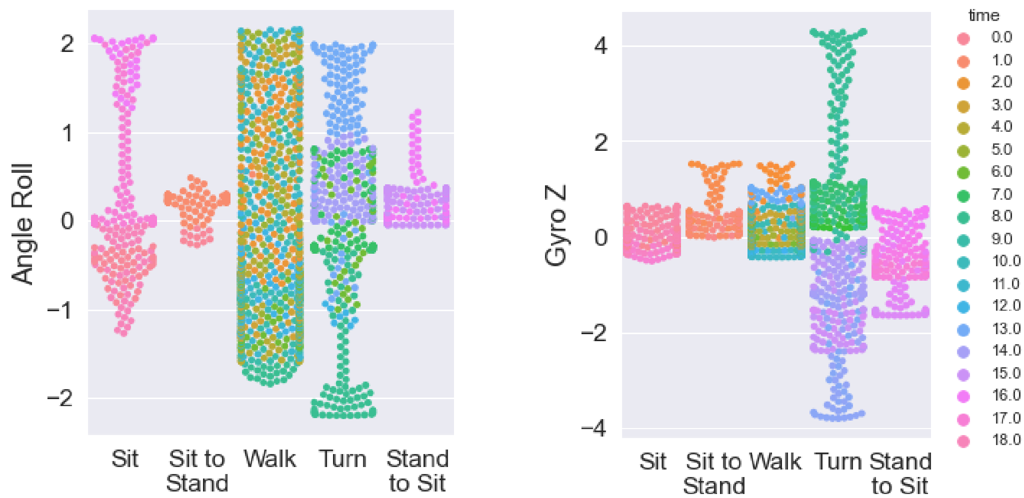
to 14th second mark, indicating that the participant started to use the support of the cane in order to sit. Similarly, the sit to stand values show that the strain increased gradually as the participant used the cane to stand up.



(a) (b)

Figure 3.3: (a) Strain Boxplot and (b) Swarm plot of strain distributed according to class labels with time as hue, time in secs, normalized by Z-score

Swarm plots of other signals like angle roll and gyroscope-z signals were plotted in Figure 3.4. Angle roll values are quite stable during the initial sitting stages, being close to a z-score of 0, and have very high variance in the walking stage as the patients' hand moves left and right. Gyro-Z values, with respect to Earth, show that the second walk stage after the first turn has a distinct concentration of negative values while there are extremely high positive values around the first turn. The peak possibly indicates the time around which the transition to turn happened (Figure 3.4). Another major observation that can be made through the swarm plots is that the characteristics of the actual transition in time from one subtask to the next could be more effective for segmentation than simply classifying the data from the various

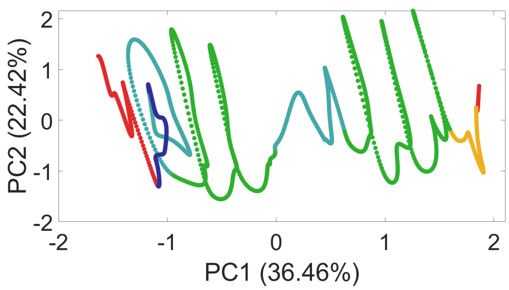


(a) (b)

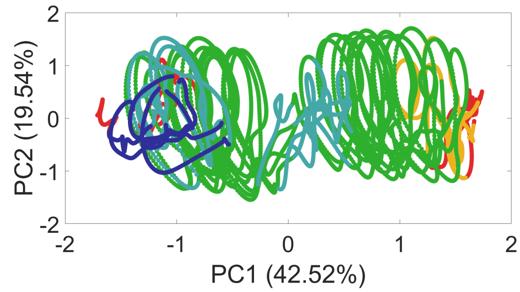
Figure 3.4: Swarm plot of (a) angle roll and (b) gyro Z values distributed according to class labels with time as hue, normalized by Z-score

subtasks as a whole.

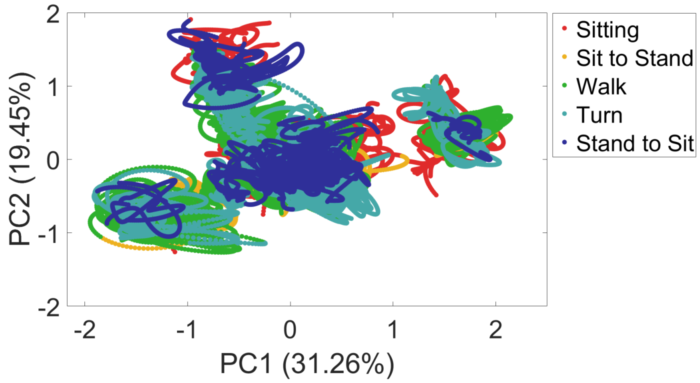
PCA plot analysis: The Principal Component Analysis (PCA) projections in Figure 3.5 show the first two principal components plotted from (a) a single trial from participant-08, (b) all trials from the same participant, (c) all trials from all 16 participants. These plots were generated by computing the first two principal components from the base signals. In a single trial PCA plot, it can be seen that stand-to-sit and sit-to-stand lie on either end, and that other subtasks also are quite separable along the x-axis. This suggests that the first principal component (which explains 36.46% variance) was quite effective in separating the subtasks. Applying PCA to all of the trials of the same participant shows that PC1 explains an even higher variance in the combined data of 42.52%. However, when combining all participants, the increased number of samples and increased inter-subject variance reduces the influence of PC1. It can be seen that the groupings of classes cluster together are representative of the temporal sequencing, suggesting that there may be benefit in trying to separate one class from the next, one at a time in a binary fashion, rather than the conventional multi-class approach.



(a) PCA - Single Trial



(b) PCA - Single participant



(c) PCA - All participants

Figure 3.5: PCA plots of (a) A single trial, (b) All trials from the participant, (c) All trials from all the participants

3.3 Training Protocols

The analysis of the subtask segmentation was conducted according to two different training and validation schemes. First, a within-person approach was explored to understand the potential performance for a given user. Next, an inter- or between-person approach was explored, which must consider differences between users. The details are as follows.

Within-person training

The within-person training approach included selecting one participant's data at a time and training a model in leave-one-trial-out fashion. In this scheme, one trial was kept for testing and the rest of the trials were used in training. This process was carried out for each trial until all trials were tested. The experiment was repeated for each participant and the average results were reported.

Between-participant

When a participant's previous data are unavailable, a model should be made generalizable enough to be tested for a completely new participant. This can be achieved through a between-participant training approach. This was done by training the system with data from all participants, excluding that of one participant, whose trials were tested on the model. This process was repeated, each time leaving out a different test participant, until all participants were tested. This training scheme is also known as a 'leave-one-person-out' approach and, although much more challenging, represents the more clinically relevant use case.

3.4 Multiclass classification - Baseline Results

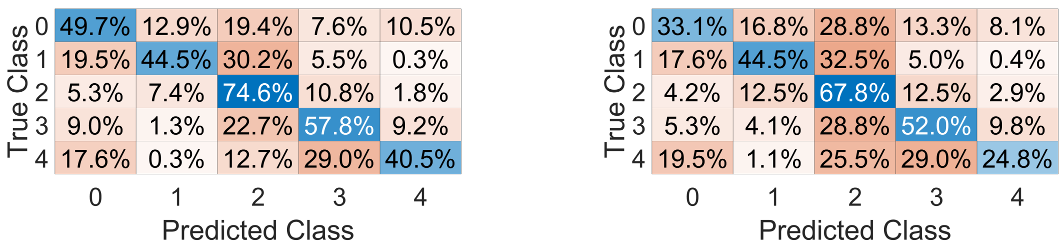
To establish a baseline approach to segmenting the different TUG subtasks, a multiclass classification method was used to predict a label at each time sample of the dataset. The inputs to the classifier were the filtered raw data, the time index, the angle pitch, roll, yaw, the accelerometer (X,Y,Z) values, gyroscope (X,Y,Z) values, and the strain. The class labels were sit (0), sit to stand (1), walk (2), turn (3), stand to sit (4). Four different classifiers were used: SVM with a Radial Basic Kernel Function (RBF), KNN with k as 1, 5, 10, 15, an LDA, and a QDA. A summary of the results is presented in Table 3.2, where the best classification results of 63.25%

were recorded using LDA for the within-person approach and 53.55% using KNN with k=15 for the between-person approach. These accuracies, however, show only how many correct samples were predicted and not necessarily how close the predictions were in time, and if the predictions even followed the order of the TUG test. The accuracies were also not particularly high, and further examination identified some underlying issues with the multiclass approach.

Table 3.2: Multiclass classification results for labelling TUG-subtasks

	Within Person		Between Person	
Model	Avg Acc(%)	STD dev	Avg Acc(%)	STD dev
KNN1	60.636	13.213	53.48585	11.87831
KNN10	60.796	13.314	53.29043	11.81297
KNN15	61.247	13.452	**53.5594**	11.96625
KNN5	60.68	13.253	53.35558	11.87716
QDA	57.237	12.575	49.43959	10.75094
LDA	**63.254**	12.916	51.43762	11.75499
SVM	61.426	12.305	52.31734	11.27375

The results were further analyzed by plotting a confusion matrix of the best and the worst performing classifiers in each case. As seen in Figure 3.6, the most predicted class is always the 'walk' class. This is due in part to class imbalances caused by walk having more samples. Thus, a classifier may learn to predict walk more often than true and still lead to higher accuracies than random guessing. This also raises a question of whether accuracy of prediction is the best metric for such a problem, as we are not trying to see if most samples got predicted correctly but we are looking to identify the transition points between classes. The conventinal multiclass approach, however, does not consider time, and thus some classes are predicted out of sequence despite the strict sequencing of the TUG test.The relatively poor accuracies seen in Table 3.2 may be partially explained by ‘sit-to-stand’, ‘stand-to-sit’ and ‘turn’ being predicted as ‘walk’ more often than they should, as can be seen in Figure 3.6. Although improved preprocessing and feature extraction approaches could lead to better performance, the mapping from classification results to actual segmentation of the test remains unclear. Furthermore, imposing more strict rules about temporal ordering could offer substantial benefits.



(a) LDA Confusion Matrix, Best Within Participant

(b) KNN15 Confusion Matrix, Best Case Between Participants

Figure 3.6: Confusion Matrices for the best cases (a) within-person, (b) Between Person classification. Note: 0-Sitting, 1-Sit to Stand, 2-Walking, 3-Turning, 4-Stand to Sit.

3.5 Summary

This chapter provided an introduction to the smart cane signals and how they can be used for classification of the different subtasks of the TUG test. A visualization of the features was used to explain how some features could be useful in segmenting subtasks. A plot of these features against each other shows some consistent patterns for particular subtasks like walking and turning. Similarly, a look at the signal plots, PCA plots, and swarm and box plots further hinted at the value of other features, such as a higher variance in gyroscope-Z data during turn, a rise in strain throughout sit-to-stand when the cane is used to stand up, and a periodicity in strain and accelerometry data during the walk subtask. This initial investigation motivates the extraction of more complex features, capable of describing the underlying patterns. Additionally, the multiclass classification results give a baseline for further evaluations that may leverage the temporal structure of the test to reduce the errors.

Chapter 4

Staged Binary Classification and Temporal Approaches

Analysis of the naive multiclass classification results in Chapter 3 motivated the notion that the temporal structure of the TUG-test could be better exploited using a staged binary classification approach. A flow diagram of the proposed process is depicted in Figure 4.1. The first step was feature extraction from overlapping windows of data to create more robust feature estimates while maintaining temporal resolution. Features were then selected using sequential forward feature selection (SFS) which was customized according to the testing procedure. To exploit the temporal sequence of events in the TUG test, a staged binary classification approach, comparing only two contiguous classes (subtasks) at a time, was implemented.

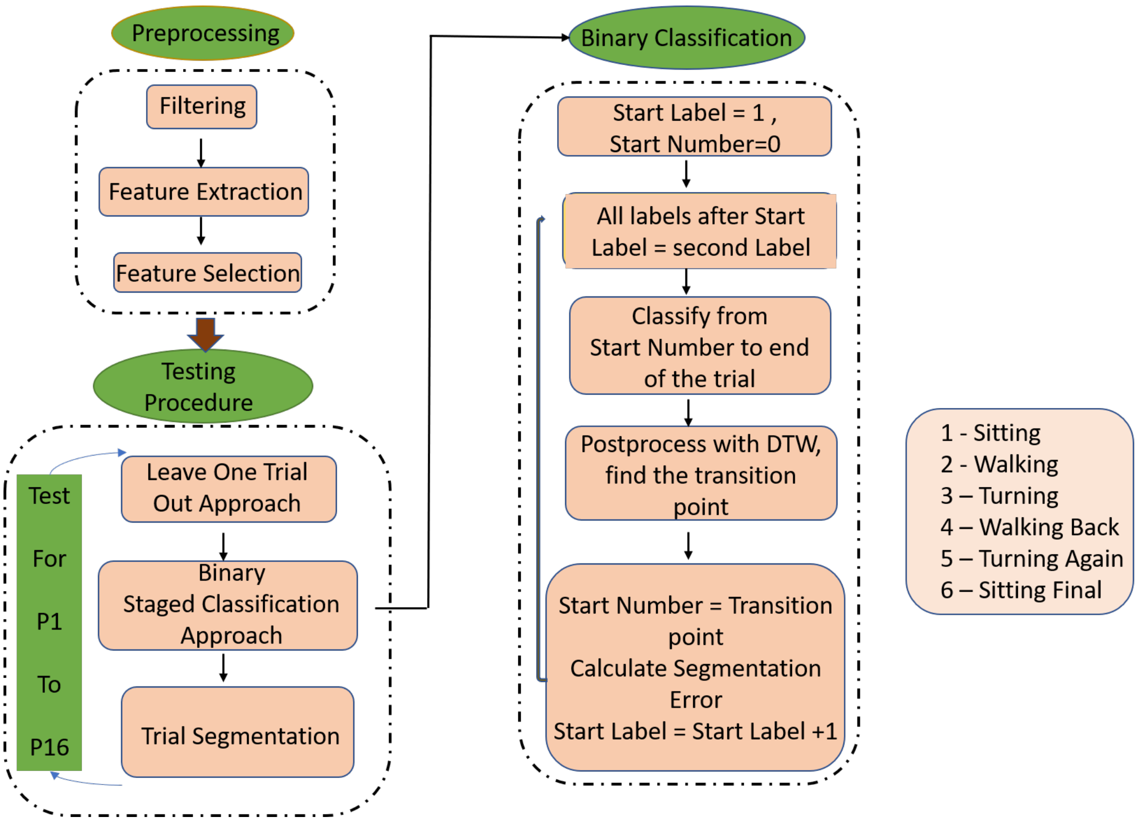


Figure 4.1: A staged binary classification approach for TUG subtasks segmentation (final architecture).

4.1 Feature Exploration and Post-processing

To better extract discriminative information from the signals, statistical and signal features were extracted from overlapping window segments of the original data. To inform the selection of appropriate window sizes, the characteristics of the lengths of each TUG subtask were analyzed and presented in Table 4.1. It can be seen that class 0 (the initial sit subtask) and class 7 (sitting after completion) had the least number of samples, with very small minimums. This was because the initiation of recording of data may have started late or was finished early, reducing the data samples in these classes in some trials. To mitigate this experimental limitation, these classes were later merged with their adjacent classes in order to improve the data distribution and the class imbalance factor. Because the length of time seated is not a clinical indicator, this was deemed to be acceptable for the purposes of this work. As before, the results again showed that the walk classes (2 and 4) were longest, and thus had the most samples, but also the most variation in time across subjects.

Table 4.1: Min, average and standard deviation of counts per class

Label	Min Count	Avg. Count	STD
0	14	186	142.46
1	64	190	118.54
2	275	516	147.67
3	89	232	94.20
4	285	476	125.65
5	20	182	88.15
6	53	175	59.13
7	12	142	285.31

The subtasks for 'turn' (label 3) and 'turn back' (label 5), however, contained few data points, limiting the potential length of window sizes that could be used for feature extraction. Window sizes above the min count for these classes would lead to missing those subtasks altogether. As an example, if a window of 100 were to be used, a total of 64 total trials (more than half of the trials) would seemingly not contain those subtask. Empirical testing with window lengths of 20-80 indicated that window sizes of more than 40 samples overly filtered the data stream, and that the best results were obtained when features were extracted using a window size of 20. These windows were extracted with a frame increment of 1 sample (10ms), to maintain the temporal resolution of the processed time series, so as to not introduce unnecessary timing error.

4.1.1 Feature Extraction

The filtered data signals were normalized using z-score standardization before extracting time and frequency domain features from the 200ms windows. Time-domain statistical features extracted from each of the original signals included the means, standard deviations (SD), maximum and minimum values and their indices, root mean squares (RMS), counts of peaks within a window, skewness, kurtosis, and interquartile ranges. Together, these yielded a (11×20) 220-dimensional feature vector. Other features extracted from each of the original signals included the total energy, and the maximum and minimum frequency powers and their indexes computed from the Fast Fourier Transform (FFT) [114], [113]. These frequency domain features therefore constituted a (5×19) 95-dimensional vector, resulting in a total of 315 statistical time and frequency domain features.

Characteristics of these features, when plotted against time, can describe differences in the subtasks by looking at the signal variations and extremes. In Figure 4.2, for example, the Interquartile Range (IQR) of the strain, which shows the difference between the quartiles in each window, shows periodicity in walking and general low values in sitting. The maximum of accelerometer-Z in each window also describes walking, with repeatable patterns during times of maximum acceleration in the Z-axis (lifting and dropping of cane) of data and a steady average while sitting (cane in steady position). Conversely, the Root Mean Square (RMS) of the angle yaw and the mean of gyro yaw indicate possible turns. By combining these features effectively through appropriate selection, it may be possible to effectively distinguish between two contiguous classes (one after the other) at a time by performing a binary classification, as will be explained later.

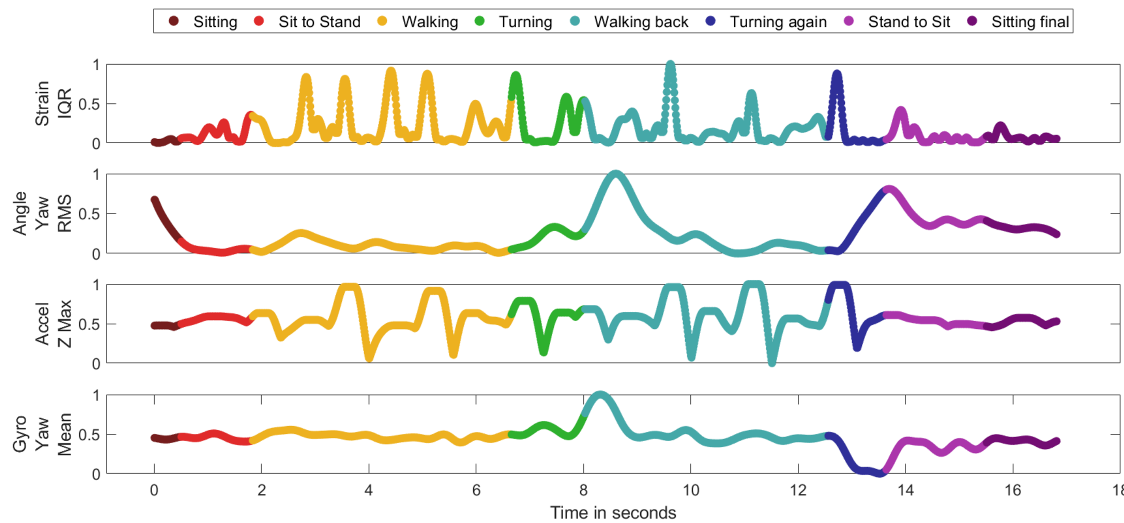


Figure 4.2: Example plots of a few features derived from the sensor data.

4.1.2 Accuracy vs. Segmentation Error

Earlier, the criterion for measuring performance was naively taken to be classification accuracy, as is common in such works. However, due to class imbalances, and the ultimate goal of segmentation, accuracy may not be the best performance evaluation metric. It would be difficult to evaluate the validity or the usefulness of the obtained results by accuracy alone, as it does not give solid information about the transition point between two TUG subtasks (errors may lead to many apparent transitions). Consequently, segmentation error (the absolute difference between the predicted and actual transition points), was adopted as the basis for performance evaluation. It was later identified that segmentation error was not only a better performance metric, but that is also constituted a better cost function (error criteria) when optimizing the binary stages of classification and during the feature selection process.

4.1.3 Dynamic Time Warping (DTW)-based Postprocessing

One problem with the output of conventional classification-based approaches to this sort of task is that they do not inherently incorporate the temporal structure of the problem - each decision is made independently of the next. The Dynamic Time warping (DTW) algorithm, conversely, is an efficient and robust algorithm used in temporal fusion [115]. By calculating point to point euclidean distances between matched sequences, DTW can be used to smooth or correct predicted classification errors, and to determine optimal transition points. This approach works well with the strict time-based sequence of subtasks seen in a TUG test as it finds the best non-decreasing aligned path. Moreover, DTW has previously been found useful in sensor based [110] and video based [69] studies in TUG segmentation. Consequently, DTW-based postprocessing was applied to the results of the classification. Briefly, DTW calculates an optimal match between two given sequences by calculating the euclidean distance between two sequences, point to point, added to the minimum of the previous adjacent path costs. The path costs are then traced to generate a matched sequence.

The optimal path is calculated using (4.1), where 'd' is the Euclidean distance.

$$D_{min}(i_k, j_k) = min_{ik-1,jk-1}\, D_{min}(i_{k-1}, j_{k-1}) + d(i_k, j_k | i_{k-1}, j_{k-1}) \tag{4.1}$$

Likewise, the DTW path cost is computed as follows.

$$D = \Sigma_k d(i_k, j_k) \tag{4.2}$$

The DTW algorithm was used as a post-processing method to smooth predictions that were out of order in a time-aligned, time-increasing manner. Using DTW, for example, as seen in Figure 4.3, a predicted sequence of a binary classification of Walk

(2) vs. Turn (3) would be compared to true labels 2 and 3 and each predicted point's distance from the base labels would be calculated using the above given equations. The lowest cost path would be traced in a reverse order starting from the last point and ending at the first, thus giving a non-decreasing path, avoiding misaligned predictions and keeping the sequence strictly increasing. This ensured that there was only a single clear transition point from one class to the next.

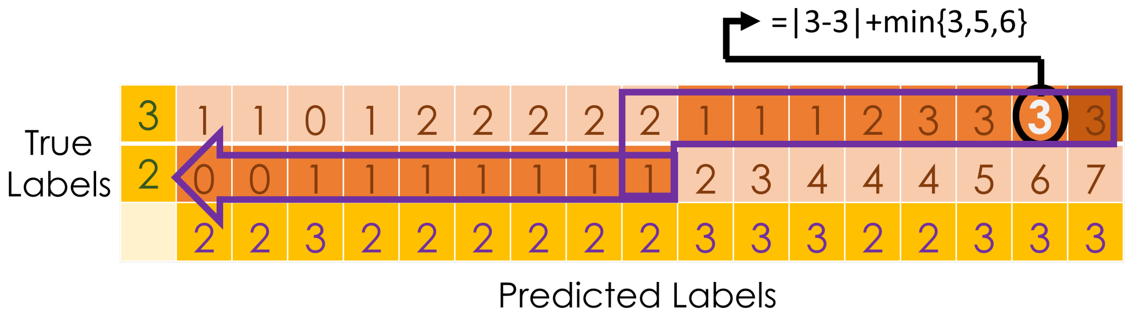


Figure 4.3: Example of DTW postprocessing on predicted labels to align and find transition point

4.2 Binary-Staged Classification

To further leverage the temporal structure of the test, a binary classification approach was applied in stages of each subtask being executed in time. That is, classification was performed by considering only two subtasks at a time, beginning with subtasks 0 vs. 1, then 1 vs. 2, and so on. Thus, 7 binary classifiers were trained in sequence to locate the TUG subtasks transition points. Initially, all of the extracted features were used to train and classify the binary subtasks sequentially, after which features for each binary classification were selected separately (as later described in Section 4.2.3). This allowed each individual classifier to be tuned to the specific feature distributions of the classes surrounding a transition.

4.2.1 Forward 1 vs. 1 binary

In this approach, all binary classifiers were trained with data corresponding to their respective two classes in both the leave-one-trial-out, and later, the leave-one-participant-out frameworks. For testing, an entire trial was classified by the first binary classifier (0 vs. 1). All the extracted 315 features were taken in the preliminary approaches. The predictions were then postprocessed using DTW to align the predictions in an ascending order in time and locate the first transition point. The segmentation error was calculated for this classifier and the transition point was then used as the start point for the next binary classifier (i.e. 1 vs. 2) to locate the next transition point. This process was repeated until all the transition points were located with the help of all binary classifiers.

Using this approach, an average total segmentation error of 13.917s was recorded as shown in Table 4.2. This metric represents the absolute sum of all the timing errors across all segments of the test, and is therefore a pessimistic, but telling, view of the results. These relatively high segmentation errors occurred because samples from later repeated classes, such as 'walk-back' and 'second turn', were often confused with earlier classes by the binary classifiers resulting in a noisy decision stream that was problematic when applying DTW postprocessing. This was particularly evident because these latter classes were not seen during training of the earlier binary classifiers (e.g. 0 vs. 1, 1 vs. 2) classes, and it was ill-determined which class they would be assigned to.

4.2.2 Forward 1 vs. all next binary

To address this disparity, a 1 vs. all next approach was evaluated, wherein each binary classifier included training data from the current class vs. all remaining classes pooled together. In this approach, for example, the 0 vs. 1 classifier included any data after subtask 1 (2-7), but with their labels changed to class 1. This was done to ensure that, during testing, the later classes were known to the classifier but associated with the 2nd class and the testing and training lengths of trials stayed proportionate. The testing process remained the same as the previous 1 vs. 1 case.

The results for the 1 vs. 1 approach, as later seen in Table 4.2, yielded higher segmentation errors, particularly when the walk class was involved (due in part to its high variability compared to the other classes). The results for the 1 vs. all next approach, however, show a clear advantage. Figure 4.4, which shows the transition estimations for an example trial and classification pipeline for participant 10, 2^{nd} trial (chosen because it represents the median of the 1 vs. 1 cases) highlights the timing differences between the 1 vs. 1 and 1 vs. next all binary cases.

Figure 4.4: Comparison between predictions of Participant (P10)-Trial 2 using 1 vs. 1 approach (Error=4.86s) and 1 vs. next all (Error=3.79s).

Merging the Initial and Final Sit Classes

Because of the aforementioned brief periods of the 'sit' portions of the experimental data, it was decided to combined them with their corresponding 'sit-to-stand' and 'stand-to-sit' subtasks [69]. This was additionally motivated by the consideration that it is the cane's movement that was being measured, and not that of the actual participant. Consequently, because some users did not use the cane to assist in standing up or sitting down, the differentiation between these two brief subtasks may have been problematic. As previously detailed, however, this was deemed acceptable as the 'sit' portion of the test holds little clinical value.

The results of merging these classes are also shown in Table 4.2. It can be seen that there is a substantial decrease in total error after merging the initial and final classes, as denoted by the bold entries. This is attributed not only to the decrease in number of subtasks but also because the subsequent binary classifications after the initial 1 vs. 2 (sit to stand vs. walk) are improved due to a more accurate initial transition.

Table 4.2: Results of the 1 vs. 1 , 1 vs. next all and merged classes approach

CLF	Method	Segmentation Error (s)							
		0 vs. 1	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	6 vs. 7	Total
SVM	1 vs. 1	0.964	2.502	3.340	3.097	1.970	1.299	0.744	13.917
	1 vs. next all	0.669	1.350	1.070	1.199	1.070	0.941	0.770	7.071
	Merged	0.743		0.915	1.210	1.070	0.941		**4.880**
LDA	1 vs. 1	1.369	2.103	3.638	4.417	4.786	3.712	3.022	23.047
	1 vs. next all	1.19	0.858	0.452	0.575	0.482	0.732	1.608	5.897
	Merged	0.817		0.793	0.834	1.040	1.300		**4.780**

4.2.3 Feature selection

Next, to optimize the feature sets for each classification task, feature selection was performed across users for each multiclass and binary segmentation case using a wrapper based sequential forward selection (SFS) approach with both SVM and LDA classifiers, separately. The models were created and tested as described above, while iteratively adding the next best feature to the feature set (based on the obtained performance). The predicted results were then postprocessed using DTW to smooth the results and find transition points, as before. The SFS was optimized using two different parameters to enable a comparison - (a) sum of average segmentation error (in seconds) and (b) accuracy (%).

Table 4.3: Average Segmentation Errors (seconds) of different models

Method	CLF	Segmentation Error (s)					
		1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	Total
Multiclass (All Feats)	SVM	0.881	0.991	0.789	0.859	0.792	3.913
	LDA	0.435	0.241	0.553	3.332	1.075	4.511
Binary (All Feats)	SVM	0.743	0.915	1.21	1.07	0.941	4.88
	LDA	0.817	0.793	0.834	1.04	1.3	4.78
Multiclass (SFS)	SVM	0.477	0.439	0.519	0.47	0.517	2.422
	LDA	0.403	0.438	0.517	0.57	0.493	2.421
Binary (SFS)	SVM	0.336	0.437	0.444	0.345	0.353	1.915
	LDA	0.346	0.48	0.411	0.309	0.352	1.898

Different features were selected by the SFS algorithm for each binary case that best described the TUG subtasks. The stark improvement between the all feature and SFS-based case seen in Table 4.3 shows that the most important features are different for each of the binary subtask classifiers. Similarly, the performance of the multiclass approach, despite feature optimization, was unable to match that of the binary classifiers. This is again likely because the optimization of every binary classifier enables the fine tuning of relevant features. Consequently, further analysis focused on the binary case.

For the LDA binary classifier, which had a total average error of 1.898s, the number of features selected for each of the 5 binary classifications were 22, 57, 27, 38 and 98, respectively. The different numbers of features selected are a result of the stopping criteria (when an increase, or no decrease, in timing error was seen by adding another feature) . Intuitively, the 'Relative time' (the percentage of time through the TUG test) was the most commonly selected feature across classifiers, as shown in Table 4.4, and was always the single best feature to discriminate the binary cases. For the first binary classifier (sitting to walking), the selection of RMS and means of the angular roll or gyro roll values is because they tended to be higher while walking than when the participant was sitting. Similarly, for the second binary case of walking vs. turning, the selection of minimum of angular yaw, gyro yaw, and maximum of acceleration yaw in the frequency domain reflects that the top of the cane turned more while the person was turning. For the third binary case of turning and then walking back, features like the maximum of angle yaw and the standard deviation of gyroscope Z values were higher during the turn and the minimum indexes of accelerometer pitch in the frequency domain and the RMS of strain were higher during the walking period. Similarly, for the walk back vs. turn back case, the minimum and RMS of gyro roll and the minimum values of the strain explained walking differences. Finally, for the turn back vs. sit case, the features included the gyro yaw and angle yaw means to indicate turns, and the minimum of strain to describe activity and inactivity of the cane.

Table 4.4: Top 5 SFS selected features for each binary classification using LDA

#CLF	#Feats	Top 5 Features
1vs2	22	Relative time, Angle roll RMS, Gyro roll angle mean, Gyro roll RMS, Acclerometer yaw min time domain
2vs3	57	Relative time, Angle yaw min time domain, Gyro yaw angle min frequency domain, Accelerometer yaw max frequency domain index, Angle yaw RMS
3vs4	27	Relative time, Angle yaw max time domain, Gyro-Z SD, accelerometer pitch min frequency domain index, Strain RMS
4vs5	38	Gyro roll min time domain, Gyro roll RMS, Angle yaw RMS, Strain min time domain, Gyro-X min frequency domain index
5vs6	98	Relative time, Gyro yaw mean, Gyro-X max time domain index, Angle yaw RMS, Strain min time domain

4.2.4 Comparison between SFS Objective Functions

In order to show the effectiveness of the proposed segmentation error-based objective function in SFS, the results were also computed using classification error (accuracy in percentage) to compare with the segmentation error. A comparison of the approaches is shown in Table 4.5, where the segmentation error-based objective function resulted in lower timing errors for both the binary and multiclass classification scenarios. In addition, the overall classification accuracy of the TUG test was also computed using both SFS objective functions. As seen in Table 4.5, the classification performance was also improved using the proposed SFS criterion in binary classification approach; however, no major improvement was recorded in multiclass classification. Nevertheless, because the improvement (or at least no decrease, in the multiclass case) was not only seen in the resulting timing error but also in the overall accuracy, the use of timing error is recommended as a superior objective function for optimization of such problems.

Table 4.5: Comparison of timing errors and accuracy based on feature selection objective function.

Classifier	Approach	Overall Timing Error (s)		Overall Accuracy (%)	
		SFS Criteria		SFS Criteria	
		Accuracy	Timing Error	Accuracy	Timing Error
SVM	Binary	2.29	1.91	90.07	91.20
LDA		2.21	**1.90**	90.00	**91.31**
SVM	Multiclass	2.43	2.42	89.84	89.19
LDA		2.29	2.42	89.35	89.56

Forward-Backward classification

Given the success of the staged binary classification approach, and the benefit of DTW-based smoothing, the process was also explored in a backward fashion, starting from class 6 vs. 5 , class 5 vs. 4, and so on (while maintaining the same SFS selected features). The point of transition for each binary classification was then recorded for both the forward and backward cases, and the average of the recorded transition was used as the final transition point decision. This was conceived as a possible way to mitigate the effect of spurious decisions that may have caused early or late transitions when only traveling in the forward direction.

These results of Table 4.6, however, reflect that the backward approach under performed the forward approach and thus their fusion did as well. In the backwards approach, the later subtasks (whose transition errors were generally higher) were evaluated first, resulting in an accumulation of errors throughout the test.

Table 4.6: Forward-Backward Classification results for SVM and LDA

#CLF	Approach	Segmentation Error (s)					
		1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	Total
SVM (SFS)	Forward	0.336	0.437	0.444	0.345	0.353	1.915
	Backward	0.487	0.414	0.496	0.656	0.465	2.520
	Average	0.417	0.426	0.472	0.433	0.410	2.158
LDA (SFS)	Forward	0.346	0.480	0.411	0.309	0.352	**1.898**
	Backward	0.463	0.428	0.629	1.520	0.642	3.680
	Average	0.392	0.447	0.501	0.836	0.476	2.652

Analysis of Outliers

The best case, as seen in Table 4.6, was with forward, binary LDA classification. This was only slightly better than the SVM results as both forward SVM and LDA had similar distributions of class segmentation errors across participants. A trial-wise look at the LDA results, as seen in Figure 4.5, however, shows that participants P04 and P06 had much higher segmentation errors than the other participants, effectively worsening the overall results. It was noted clinically that P06, in particular, used the cane very differently (almost sideways) due to a severe pathology.

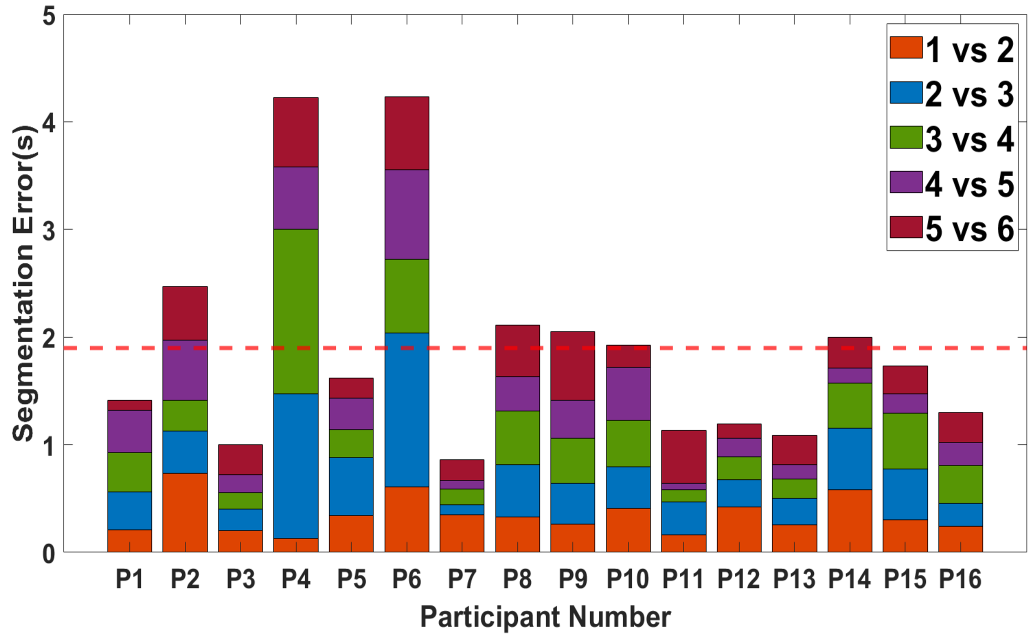


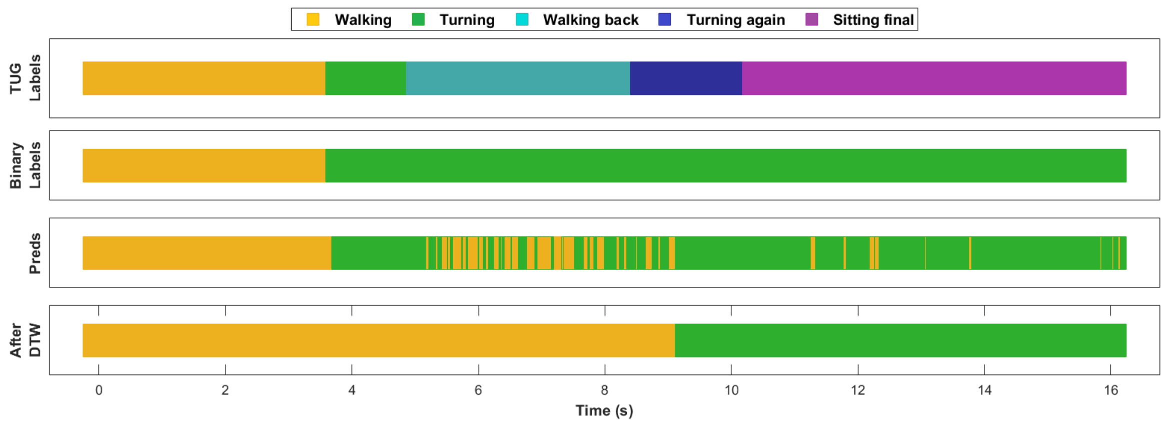
Figure 4.5: Participant-wise average class segmentation errors

Looking at P04, a single outlier trial contributed to the larger segmentation errors (see Table 4.7). Investigating that trial more closely revealed prediction errors in 2 vs. 3 (walk vs. turn) due to an apparent partial turn during the walk phase. This was smoothed after DTW correction, but led to the shift in the predicted transition point (Figure 4.6).

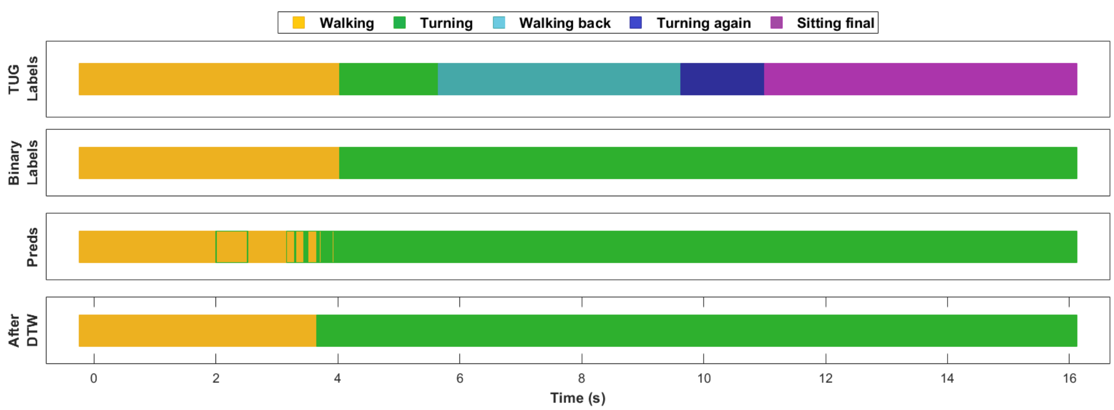
Table 4.7: Per trial error analysis of the participants having largest segmentation errors. Right side: Description of the errors occurred in different TUG subtasks in most erroneous trial

Trial	Total Error (seconds)	
	Participant (P04)	Participant (P06)
T1	0.98	4.34
T2	**16.5**	4.81
T3	0.64	1.62
T4	1.9	6.03
T5	1.23	4.35

Participant (P04)	
Trial (T2)	Error
1vs2	0.20
2vs3	**5.52**
3vs4	5.84
4vs5	2.31
5vs6	2.58



(a)



(b)

Figure 4.6: Analysis of binary (2 vs. 3) segmentation errors using LDA: (a) Trial 2 (P04), where DTW-based post-processing misaligned the LDA predictions, (b) Trial 3 (P04), where DTW corrected the erroneous predictions and reduced the segmentation error.

4.3 Between-Person Segmentation

To evaluate the potential performance of the approach across subjects (subject independent), the same optimization methods were applied in a leave-one-person-out approach. Here, each trial from a given participant was tested, but using training data taken only from the other participants. This framework gives a better view of the generalized results without requiring a custom model to be trained and built for each user. The results in Table 4.8 show that the errors are, predictably, higher than the within-person approach. Nevertheless, despite very visible differences in the way the cane was used by different participants, differences in relative timing of subtasks across participants, and some participants having fewer trials than other, the results remain promising.

Table 4.8: Between-participant TUG subtask segmentation errors

#Feats	#CLF	Segmentation Error (s)					
		1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	Total
All	SVM	0.7432	0.915	1.2069	1.0718	0.9407	4.8776
	LDA	0.7761	0.6395	0.7768	0.5892	0.7402	3.5218
SFS	SVM	0.7771	0.8252	0.8393	0.5191	0.8145	3.7752
	LDA	0.69609	0.67083	0.70007	0.50784	0.84193	3.4167

4.4 Summary

In this chapter, results were described for the within- and between- participant testing cases. In both the cases, LDA outperformed SVM by a small margin. SVM focuses on the classification of support vectors which are the points closest to the decision boundary in a feature space. These, however, may not necessarily be points on the decision boundary for a temporal measure, as in this work, where points in time space between two classes are the important points to classify. LDA tends to linearly maximize the separability of all considered points, which may explain why LDA slightly outperformed SVM in each case. The proposed staged binary classification process leveraged the temporal nature of the TUG test, and enabled features to be tailored based on binary cases. Finally, the use of segmentation error as the optimization criteria improved both segmentation error and accuracy results, as compared to the conventionally used accuracy.

Chapter 5

Extended generative and deep learning approaches

5.1 Hidden Markov Models (HMMs)

HMMs are popular probabilistic models that provide a solution to three common machine learning problems - evaluation, decoding and learning - to find the highest likelihood for classification [116]. Importantly, they are a special case of Bayesian networks and inherently incorporate the type of temporal awareness that was found to be beneficial in the previous chapter. Consequently, they were explored here as a potential improvement over the previous staged binary classification pipeline. They work by modeling not only the specific classes, or states of the series, but also the probabilities of transitions between them. Briefly, for a HMM model λ and observation sequence, O (e.g. $O = O_1, O_2..O_T$), assumed to be generated by a state sequence $(Q_1, Q_2...Q_T)$ of length T, the model can be defined by the number of states (N), the number of observations (M), the initial probabilities (π) of being in a state, the transition probabilities ($A = [a_{ij}]; i, j = 1, 2..N$), where a_{ij} is the transition probability from state i to state j and the observation probability $b_j(O_k)$ which is modeled with the continuous output probability density function [117, 118, 119]. The density function is denoted by $b_j(x)$, where x represents the k dimensional feature vector. Next, the observation probability or likelihood can be calculated using O, Q and π values.

To apply HMMs in the TUG segmentation context, the model was trained sequentially by feeding the feature data, consisting of the six TUG segments as classes, along with their corresponding labels. The Baum-Welch algorithm [120] was used to maximize the likelihood of estimation of the hmm parameters from the training set by re-estimating the initial output probability distributions. The Viterbi decoding algorithm was then used to identify the most probable sequence of subtasks for a particular input sequence of features [121].

It was initially believed that the HMM approach could be used as a multiclass solution, modeling the entire TUG as one sequence. However, initial empirical testing yielded poor overall segmentation errors (see Table 5.1) when combining all subtasks together. This is likely due to the limited number of training trials from which to learn all of the parameters of state transitions probabilities needed to explain the entire TUG sequence. Consequently, a staged binary classification approach was again adopted to break the learning problem down into small segments. As before, 5 binary HMM models (i.e. 1vs2, 2vs3, etc.) were trained for TUG-subtask segmentation. To do this, observation sequences of 40 samples of each of the raw signals were used. The HMMs were initially trained and tested by varying the number of allowed states in each model from 2 to 7, and the number of Gaussian mixtures used to define the feature distributions within each state from 1 to 128. Although the classifiers were binary, additional states were allowed to potentially model each subtask as a sequence of states (eg. the periodic nature of gait could be described by a repeating sequence of states, within that single subtask). The application of the binary staged HMMs was then done as before, and DTW-based postprocessing was applied to the resulting decision sequence. This approach yielded an average total timing error of 5.271s.

Table 5.1: Results for Multiclass and Binary HMM within-person segmentation using the raw sensor values

Method	Segmentation Error (s)					
	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	Total
Multiclass HMM	5.0312	2.135	4.477	1.930	1.493	15.067
Binary HMM	1.008	0.900	1.106	1.215	1.042	**5.271**

To improve these results, the 315 time- and frequency-based features were extracted as explained in Chapter 4 maintaining the labelled data information. The staged binary HMM approach was then repeated using the extracted features, yielding the best combinations of states and Gaussian mixtures shown in Table 5.2. The high numbers of mixtures required to explain the individual states speaks to the dynamic nature of the test, and may have been worsened by the high dimensionality incurred by using all features.

Table 5.2: Number of HMM states and Gaussian mixtures determined for within-person TUG-subtask segmentation using the full set of 315 features

#	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6
HMM states	5	4	5	4	5
GMM numbers	80	40	16	128	48

HMM Feature Selection Given the challenges with dimensionality and limited trials, a minimum redundancy maximum relevance (mRMR) feature selection technique was used to further improve the HMM results [122]. SFS could not reasonably be employed as before, because of the computational burden of learning and testing the HMM models for every required combination of features. By contrast, mRMR is a filter-based feature selection technique used to select features based on a high correlation with the provided class labels while maintaining low correlation between the selected feature vectors. The correlation with the class labels (relevance) is calculated with the F-Statistic while the Pearson correlation coefficient estimates the correlation between features (redundancy). Features are then selected sequentially by a greedy search to maximize the relevance while reducing redundancy. MID (Mutual Information Difference criterion) and MIQ (Mutual Information Quotient criterion) are commonly used as objective functions. A different number of features was then selected for each of the binary staged classifiers (225, 210, 150, 150, 175 respectively) based on empirical testing. The results of this testing can be seen in Figure 5.1, which shows the impact of number of features for each binary classifier. The segmentation errors based on this staged binary HMM classification using the mRMR-determined features are presented in Table 5.3, for both the within- and between- person cases. The use of the mRMR feature selection technique improved the segmentation results (by 1.485 s in the within person case when compared to Table 5.1) while also substantially reducing the computation time.

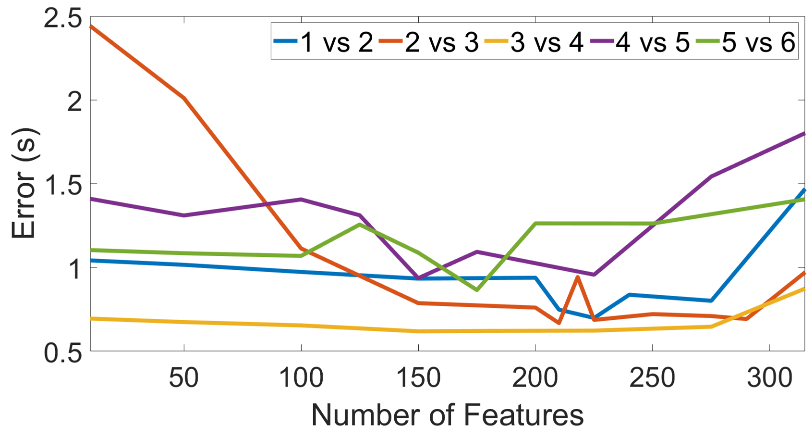


Figure 5.1: Segmentation errors for the various TUG-subtasks based on number of features. Note - 1-Sitting, 2-Walking, 3-Turning, 4-Walking back, 5-Turning again, 6-Sitting Final.

Table 5.3: Segmentation errors using staged binary HMM classification with mRMR-selected feature sets

Method	Segmentation Error (s)					
	1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6	Total
Within Person	0.698	0.668	0.619	0.936	0.865	**3.787**
Between Person	0.993	0.838	1.167	1.171	1.306	**5.475**

The mRMR-based staged binary HMM results for each participant (shown in

Figure 5.2) show a similar trend as in Chapter 4, wherein Participants 4 and 6 were substantially worse than other participants. Not only does this raise the overall perceived error, but it also contributed to similar trends in the between-person approach. Although the individual segment errors for the between-person case were all under 1.3s, the highest errors were found between the second turning and sitting subtasks, suggesting higher variation (different strategies) in the movement and loading of the cane during the final turn and sit transfers.

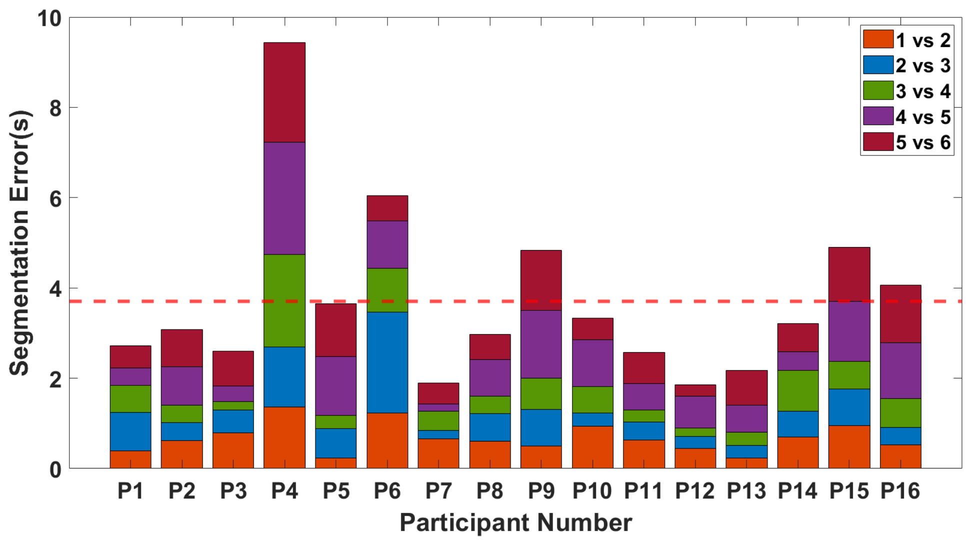


Figure 5.2: Person-wise segmentation errors in different TUG-subtasks using HMM. Note - 1-Sitting, 2-Walking, 3-Turning, 4-Walking back, 5-Turning again, 6-Sitting Final.

5.2 Feature Learning-based Methods

It is clear from the above results that some features provided more value than others, but that many features were needed to explain the subtasks. As an alternative to conventional feature extraction, two different deep-learning architectures, i.e., Convolutional-LSTMs (CLSTMs) and Encoder Decoder-Temporal Convolutional Networks (EDTCNs), were explored. Both of these techniques have been widely explored for action/activity segmentation from video/sensor data, and leverage the temporal aspects of the signal [3]. As before, the learned models were used to predict a decision stream/sequence for the test, which was then post-processed using DTW, as seen in the previous chapters.

Convolutional Neural Networks (CNNs) & LSTMs (CLSTM)

CNNs are well known for automatic feature extraction, learning to directly extract them from raw data [123]. A CNN is composed of an input layer with input data, followed by a convolutional layer which is used for feature mapping, usually using Rectified Linear Unit (ReLU) filters. The input features are then passed on to a max pooling layer to reduce the number of dimensions and thus the computational expense. This data is then passed to a fully connected classification layer with weights, biases, and neurons that ends with a logistic (binary) or softmax (multiclass) layer to produce the final output. In this work, 1D CNNs were used to process the cane data spatially as they are applicable to problems with limited labelled data and high signal variation [124]. Also, they have reduced computational complexity by involving calculations on 1D arrays rather than 2D matrices.

LSTMs are a special kind of recurrent neural networks which are capable of learning long-term dependencies. The LSTM unit contains a cell responsible for remembering values over arbitrary time intervals with the help of three gates (input, output and forget) which regulate the flow of information. These networks are highly useful in time series classifications such as event detection in gait [125], and human activity recognition [126, 127].

A CLSTM architecture was formed in this work by combining CNN and LSTM together to build a spatio-temporal model. The CLSTM network used a 1D CNN to extract spatial features from the base signals as the first step. These feature maps were later processed temporally at higher layers using the LSTMs. The LSTM implemented in this work was a many-to-many LSTM network which maps from an input sequence to an output sequence by computing the network unit activations iteratively [128]. Different combinations of window sizes (20, 40, 76, 104, 132 samples) were evaluated to find the best segmentation performance. Likewise, different convolution filter sizes (5, 10, 15, 20) were explored in each of the CNN layers. The network was trained with an Adam optimizer to reduce cross entropy loss. For classification, the data was divided into training, validation and test sets. The performance was evaluated in leave-one-trial out fashion where one trial was kept for testing, one for validation (necessary to improve generalization because of the learning process of these networks), and the rest for training. The process was repeated until each trial was used for testing.

Encoder Decoder-Temporal Convolutional Network (ED-TCN)

The ED-TCN architecture was recently proposed, [3], for action segmentation in video data. The architecture combines concepts from temporal CNNs and encoder-decoder architectures, and thus consists of a hierarchy of temporal convolutions, pooling, and upsampling layers that are able to capture long-range temporal patterns. The Encoder-Decoder architecture [129] uses neural networks both as the encoder and decoder. It takes the signals as input and encodes them into a context value in a latent space, before then decoding that context value into the expected output sequence. In this way, the ED architecture compresses the feature or signal information contained in the whole input sequences into a fixed-length vector to make tensor flow more stable [130]. Temporal Convolutional Networks (TCN) are based on causal convolutions and predict labels based on the past and present inputs. Studies, such as the one by Bai et al. [131], have found that TCN may outperform LSTM for some sequence modeling tasks. Consequently, a combined Encoder-Decoder Temporal Convolutional Network (ED-TCN) [3], whose framework can be seen in Figure 5.3, was adopted in this work.

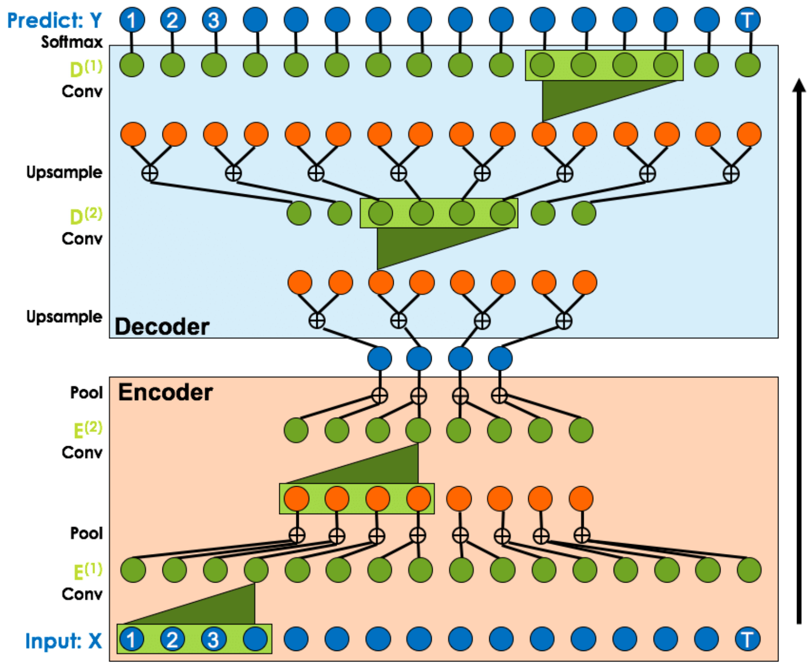


Figure 5.3: ED-TCN model using temporal convolutions, pooling, and upsampling layers [3].

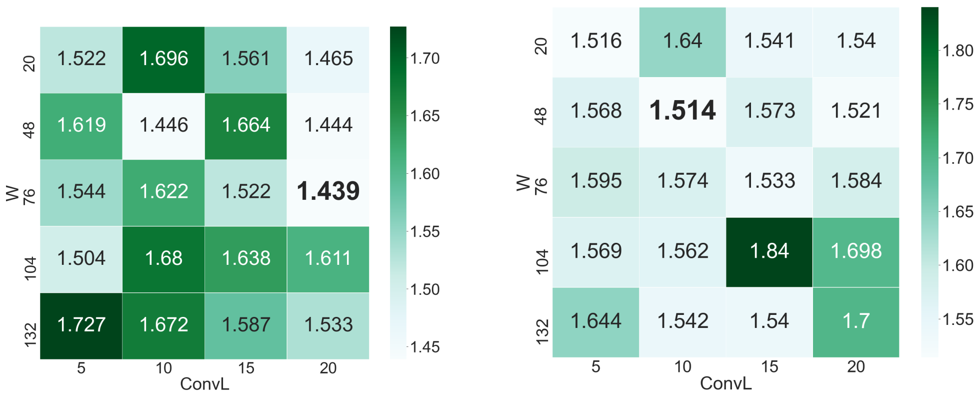
Both the encoder and decoder included two layers, with an input composed of each of the binary classes passed to the encoder. The model was chosen to be causal since the prediction at time t depended on the predictions leading up to time t-1. Hence, the input layer was zero padded at the beginning by half the convolutional filter length to avoid loss of samples before applying a 1D convolutional layer with a dimensionality output of 64 (number of kernels). This layer created a convolution kernel that was convolved with the input layer over a single temporal dimension, producing a tensor of outputs. This was then cropped at the end by half the convolutional filter length and a 1D spatial dropout was applied as the adjacent frames within the feature maps were strongly correlated. This was followed by a channel normalized ReLU activation, and finally a max pooling of 2 widows. The above steps were again implemented for a dimensionality output of 96 on the encoder side. The decoder followed the same steps, except that it included an upsampling with a factor of 2 applied to begin. As with the CLSTM, the network was trained and tested for different window sizes ranging from 20, 48, 76, 104 and 132. The convolution filter length was also varied from 5 to 20 with an increment of 5. The fully connected output layer had a time distributed wrapper consisting of a dense layer with a softmax activation to enable a probabilistic interpretation of the outputs. The most likely outputs were selected as the predictions and then passed through DTW post-processing as before for alignment and calculation of transition points.

5.3 Network Results

5.3.1 Within-Person Results

Both the proposed staged binary and multiclass approaches were investigated for each of the networks. Both CLSTM and ED-TCN test results were calculated based on the best parameters (window-size & conv filter) found on the validation sets. Heat maps of the average total TUG subtask segmentation errors based on different window and filter sizes for the binary CLSTM and ED-TCN models are shown in Figure 5.4 . It should be noted that these are the performances on the validation sets, as they are the only sets used to learn parameters, and not the final test results. From these, it can be seen that the best window size and convolutional filter length for the binary CLSTM were 76 and 20, and 48 and 10 for the EDTCN. The best window size and convolutional filter length for the multiclass case were 104 and 20 for CLSTM and 48 and 15 for EDTCN.

Based on the best parameters given by the validation set, results were computed for the test trials, yielding an average binary segmentation error of 4.354s for the ED-TCN and 4.638s for the CLSTM (See Table 5.4). Although these results indicate that the networks were indeed able to learn valuable information that helped in segmenting the tests, further improvement is needed. As with previous multiclass attempts, the models struggled to learn when exposed to all subtasks at once. It



(a) CLSTM (b) EDTCN

Figure 5.4: Heatmap of showing TUG-segmentation errors across participants for various time-windows and conv filter length in binary classification: (a) CLSTM (b) EDTCN.

is likely that the reduced performance as compared to the previous results obtained by the conventional classifiers, can be attributed to the relatively small number of samples available for training these complex networks (4 trials, given that 2 were needed for validation and testing). As seen in the Table 5.5, the binary errors were quite low (1.439s for CLSTM and 1.514s for EDTCN) during the validation process, but this did not generalize sufficiently to the test set.

Table 5.4: Within-person stage binary and multiclass TUG-subtask segmentation errors using CLSTM and EDTCN for validation (valid) and testing

Classifier	Result	W	Conv	Segmentation Error (s)					
				1vs2	2vs3	3vs4	4vs5	5vs6	Total
multiCLSTM	Valid	104	20	0.549	0.360	0.375	0.494	0.481	2.260
	Test			1.411	1.750	1.741	2.016	2.127	**9.045**
multiEDTCN	Valid	48	15	0.682	0.450	0.493	0.671	0.627	2.922
	Test			1.436	1.622	1.686	2.050	2.315	9.109
binCLSTM	Valid	76	20	0.254	0.223	0.320	0.334	0.309	1.439
	Test			0.888	0.934	1.003	1.043	0.769	4.638
binEDTCN	Valid	48	10	0.323	0.200	0.357	0.333	0.301	1.514
	Test			0.968	0.787	0.968	0.902	0.729	**4.354**

The person-wise distribution of errors for the binary classification approach, shown in Figure 5.5, show that performance on participants 4, 6, and 10 was worse than the others. As noted previously, P06 used the cane in a anomalous way, and P04 again had one trial that heavily skewed the results (total error of 20.47s for that one trial). Interestingly, P10 had high variance in the trial length, suggesting that the networks struggled with this variability, potentially due to insufficient training examples. The distribution of colors in Figure 5.5 also shows that the highest recorded errors were with the walk vs. turn class.

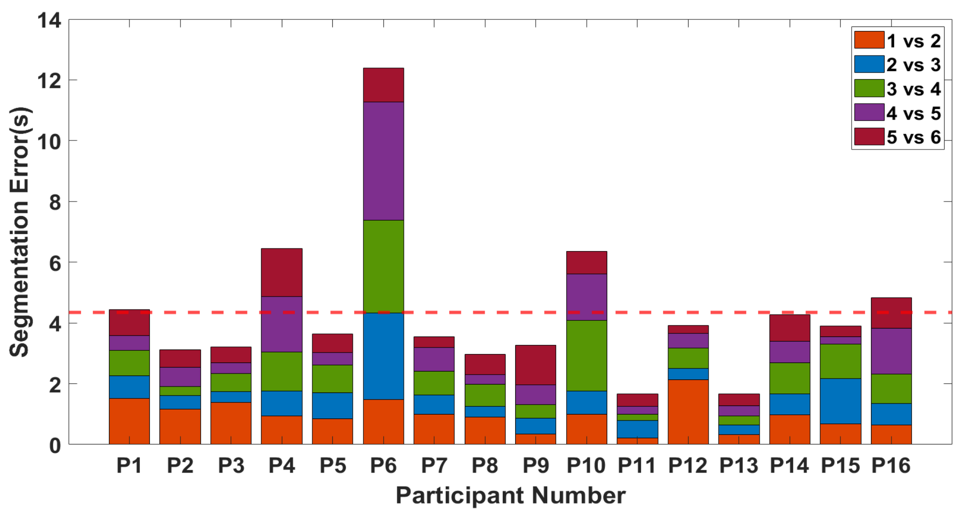


Figure 5.5: Person-wise stage binary TUG-subtask segmentation errors using EDTCN.

5.3.2 Between-Person Staged Binary Deep Learning Results

As before, the same deep methods were explored for the between-person approach. Given the data demands of the deep networks, it was thought that the availability of additional data due to the inclusion of many subjects during training may benefit them. However, with more data also came the increase in inter-subject variability, leading to higher errors across all window size and convolutional filter length combinations for both EDTCN and CLSTM, as shown in Figure 5.5. The person-wise binary EDTCN results shown in Table 5.6 reveal a sharp increase and high variability in errors, suggesting that the data from some subjects dominated the learning process. On closer inspection, it can be seen that for the participants with very high errors, the segmentation generally couldn't predict the first binary case of sit vs. walk effectively, leading to compounded error throughout the subsequent cases of binary classification.

Table 5.5: Between-person results for CLSTM and EDTCN

#CLF	W	ConvL	Segmentation Error (s)					
			1vs2	2vs3	3vs4	4vs5	5vs6	Total
CLSTM	76	20	1.964	3.768	3.743	0.921	0.733	**11.130**
EDTCN	20	20	1.671	3.909	4.181	1.369	0.595	11.724

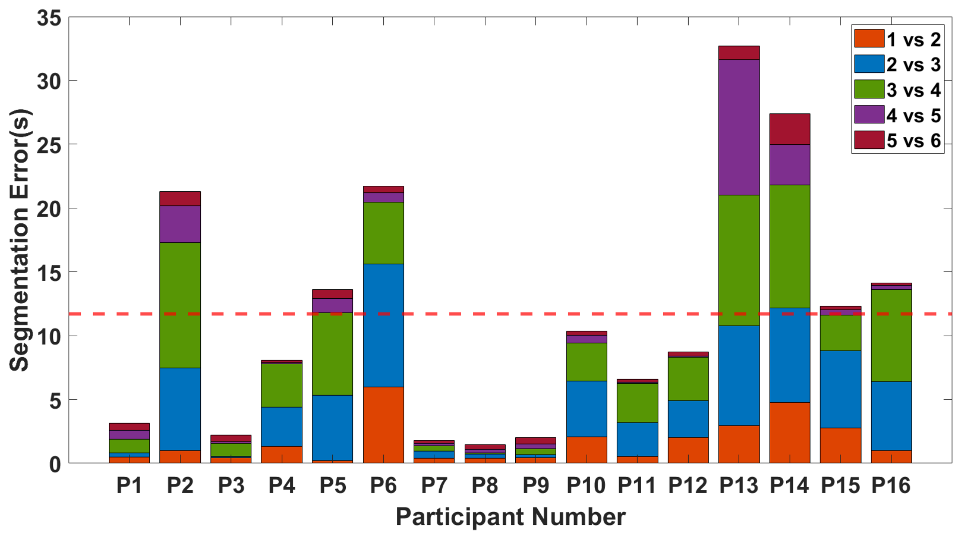


Figure 5.6: Personwise Forward best case validation results between person

5.4 Summary

In this chapter, extended approaches were explored as potential options for segmenting the TUG test. A generative approach with HMM models was tested, followed by automatic feature learning with two deep learning models - CLSTM and ED-TCN. The testing for these generative and deep learning models was done using both within-person leave-one-trial-out, and between-person leave-one-person-out approaches. It was observed that the best results overall for the within-person approach were obtained from the HMM models with an average error of 3.787s. For the deep learning results, although the results for a validation set were promising, with average errors of 1.439s and 1.514 for CLSTM and ED-TCN respectively, the results of unseen test trials did not generalize as well, with higher averages of 4.638s and 4.354s, respectively. With the HMM, the results of the staged binary approach again outperformed the multiclass approach, this time by over 65%. Similarly, for EDTCN and CLSTM, they improved by 53.14% and 47.28%, showing that the binary segmentation method is consistently superior across algorithms. Importantly, both the generative and deep learning approaches employed in this chapter showed promise, but could be improved further with the availability of more participants and more trials per participant, to enable more robust estimation of transition probabilities, mixture parameters, and network weights.

Chapter 6

General Discussion

6.1 Contributions

This work was motivated by the need for automation of clinical tests of mobility, so that they may eventually be conducted in the community. Specifically, it aimed to segment (in support of automating) the assessment of a cane assisted TUG test to eliminate the need for manual intervention or laboratory controlled environments. This is expected to improve repeatability, save time on testing, provide physicians with valuable patient data without the need for visits, and enable long term monitoring for proactive healthcare. Automated and pervasive gait monitoring will enable continued assessment of stability, strength, speed, and balance, reducing the need for unnecessary interventions, while making regular testing more feasible and adaptable to a person's needs. Through this work, it is anticipated that the analytical measurement of gait parameters and quantitative assessment of rehabilitation, fall risk, and mobility will be improved.

In this thesis, a novel approach to segmenting the TUG test was proposed, using an instrumented cane for individuals with musculoskeletal injuries or related gait disorders. For a within-person framework, an average segmentation error of 1.9s was obtained using a LDA-based approach with an error of less than 0.48s for each subtask. Transition-specific feature selections, leveraging a staged binary classification approach and based on optimizing segmentation errors, yielded improved performance of over 60%, compared to conventional classification frameworks. This suggests that carefully tailoring feature sets to classify between each pair of activities is beneficial. The use of a binary segmentation approach with DTW postprocessing to find the transition point outperformed the standard multi-class classification scheme, as it better leverages the temporal structure of the TUG test. In this work, results were evaluated and presented based on the total segmentation timing error. This approach also yielded strong average accuracies such as with binary SFS-based LDA, which was 91.3% for the within-participant results, and 85.0% for between-participant results. These compare favorably with previously reported TUG segmentation accuracies, such as the 81.8% between-participant results in [113], based on shoe mounted inertial measurement units with Parkinson's patients. This improvement is especially notable, considering that the segmentation here was performed on the cane data and not using data recorded from the actual participant using inertial sensors or cameras. When compared to video-based TUG segmentation, the 1.9s and 3.4s errors (for within- and between- participant, respectively) also outperform the 4.1s reported in [69], which used Openpose + LSTM for framewise classification and then DTW for subtask segmentation. The use of an instrumented cane, for patients who already use an assistive device, would also provide added benefits for a physiotherapist by providing loading and balance information not available from video or wearable-based approaches.

The best average total subtask segmentation error for the within-person classification was 1.9s, with an average classification accuracy of 91.3% using LDA. It is important to note that this metric denotes the total error across all subtasks. Considering only the average error in timing between the first and last TUG subtasks, which is more directly comparable to the conventional timing of the full test as a whole, the average error was found to be only 0.201s. These results, presented in Table 6.1, show that, even for the worst performing subjects, the total TUG timing error was 0.5s. This error was calculated as the absolute difference in time between the time from starting to walk to starting to sit, based on the proposed LDA model predictions and the actual TUG test labels. This error is well within the standard error of measurement (SEM) for clinical use of the TUG, which is between 0.31s and 0.96s (0.86s for using a stopwatch with a 6m TUG test) and whose minimum detectable difference ranges from 0.86s to 2.66s [132]. This suggests that the results obtained in this thesis may be sufficient for clinically meaningful automation of the TUG using an instrumented cane. This will not only help in the unsupervised collection of clinical data, but also provide the additional benefits of segmentation, along with potential benefit of specific features related to speed, balance, and strain.

Table 6.1: Participant-wise actual TUG times, subtask classification accuracies and timing errors, and total TUG time predictions, for the within-person approach

#	True Mean TUG Length (s)	True STD of TUG (s)	Subtask Class Acc (%)	Total Subtask Timing Error (s)	Total TUG Timing Error (s)
P01	16.647	0.404	92.629	1.412	0.190
P02	21.740	2.668	90.445	2.474	0.370
P03	18.404	0.693	95.323	1.000	0.040
P04	15.238	0.862	85.901	4.223	0.522
P05	13.570	0.377	88.496	1.620	0.006
P06	25.466	4.456	86.536	4.235	0.370
P07	13.097	0.750	94.153	0.858	0.285
P08	15.524	0.915	87.963	2.108	0.004
P09	19.546	0.568	91.194	2.050	0.104
P10	18.378	3.094	91.757	1.922	0.282
P11	17.095	1.520	93.915	1.137	0.117
P12	18.730	0.499	94.507	1.193	0.100
P13	17.370	0.776	94.499	1.085	0.150
P14	28.572	1.728	93.909	2.008	0.473
P15	11.120	0.632	87.384	1.732	0.070
P16	14.234	0.706	92.319	1.303	0.138
AVG	**17.796**	**1.291**	**91.308**	**1.898**	**0.201**

Furthermore, this can lead to extracting pseudo-TUG tests from daily activities, providing a larger amount of collected data in a home/community dwelling setup for a patient to enable prolonged monitoring and active prevention of mobility, balance, and fall related issues.

6.2 Limitations

While using the cane data, the two activities ('Sit' & 'Sit-to-Stand') in the beginning and two activities ('Stand-to-Sit' & 'Sit-Final') in the end were merged. This was done because of experimental conditions that occasionally caused very short sitting periods, because 'Sit-to-Stand' & 'Stand-to-Sit' had very few samples in some cases, and because some participants did not use the cane while standing up or sitting down. However, these could potentially be further decomposed using a heuristic or binary classification approach in future works.

Although the within-person results presented here yielded errors that are within the reported minimum detectable difference for the TUG test (suggesting that this may well be a clinically viable approach), it requires previously labeled data from each user. This detracts from the clinical utility of the approach, and may not be feasible in all cases. The between-person results, however, are also promising (with average subtask errors of 0.683s), and could still be improved with more participants and higher numbers of trials per participant.

Some of the participants were unable to complete the post- trials of the TUG test, leaving them with only 3 trials of the TUG test available for use in this study. This was a major setback for within-participant testing, particularly in the deep learning and temporal methods. Deep learning and temporal architectures require more data for training and for generalizability of results. In these participants, high variations were recorded in TUG timings and in their subtasks due to different speeds, swing of the cane, and strain being put on the cane. In order to improve the model with these variations in the data, it would be beneficial to record more repeated trials from each participant. This would also aid in better training the between-person models, and avoid potentially biasing them with different amounts of training data. As was shown, the conventional classifiers seemed to work better than the extended HMM and deep learning approaches. This makes sense, given the amount of training data available, as they have fewer hyperparameters to tune and are less prone to overfitting in the case of fewer trials.

It can also be seen from Table 6.1 that participant P06 had the highest standard deviation in their TUG times while taking one of the highest average time to complete a TUG test. This was because P06 struggled with standing up, sitting down and turning, and had high deviation in their subtasks. Some errors were also introduced in certain trials because the participant did not always start walking after standing, and did not sit down immediately after turning at the end. At one point, one participant (P07) even changed the cane from their left to right hand during the test. The results reported here may therefore have been improved with additional pre-processing to filter out anomalous trials or participants, or those that had insufficient data to be characterized.

Finally, although the performance of the HMM and deep learning models presented in Chapter 5 did not surpass that of the proposed staged binary classification with DTW, their results are promising. It is known that these models require substantially more data in order to learn and generalize. This was known before testing them; however, their omission from this work would have been notable. As a result, their implementation and testing here may be seen as a preliminary investigation as motivation for future works with more data. Despite this, the methods outlined, and the data collected from clinical patients, remain a meaningful contribution of this thesis.

6.3 Future Work

This work lays a foundation for the automation and segmentation of a TUG test using instrumented assistive devices. This would not only provide benefits for remote and unsupervised monitoring, but also for objective longitudinal studies. A potential approach to reducing the inter-subject variability, improving the between-participant approach, would be to adopt transfer learning. Transfer learning would enable the use of a pre-built model trained with other users (or even other tasks), reducing the amount of data needed to train a given model. This would offer the benefit of better initial predictions, improvement in best predictions, and a higher level of convergence due to availability of abundant data from which the related model has been built.

The availability of large databases of relevant data has been critical for the advancement of fields like gait analysis. Similarly, here, it would be beneficial to have more data, including more participants and trials per participant, as they may provide various other scenarios where activities were performed with differing times or parameters (strain, speed, angle shifts, etc). Continued learning about specific subtasks could further lead to automating the extraction of specific TUG subtasks from regular activities of daily living such as getting up from a chair, fetching something and coming back to sit down. These data could be compiled as part of pseudo-TUG tests and used for assessment. Finally, the temporal methods applied in this work to the instrumented cane may also be similarly applied to other modalities such as video cameras, pressure tiles, wearables, etc. Through this work, it is hoped that preventive gait and mobility monitoring may one day be facilitated, enabling better rehabilitation and early detection of mobility and balance issues.

Bibliography

[1] W.-H. Wang, P.-C. Chung, G.-L. Yang, C.-W. Lin, Y.-L. Hsu, and M.-C. Pai, "An inertial sensor based balance and gait analysis system," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, July 2015, pp. 2636–2639, doi: 10.1109/ISCAS.2015.7169227.

[2] S. Gill, J. Hearn, G. Powell, and E. Scheme, "Design of a multi-sensor IoT-enabled assistive device for discrete and deployable gait monitoring," in *2017 IEEE Healthcare Innovations and Point of Care Technologies, HI-POCT 2017*, Bethesda, MD, Nov 2017, pp. 216–220, doi: 10.1109/HIC.2017.8227623.

[3] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," *CoRR*, vol. abs/1608.08242, pp. 47–54, Aug 2016. [Online]. Available: http://arxiv.org/abs/1608.08242

[4] S. Morris, M. E. Morris, and R. Iansek, "Reliability of Measurements Obtained With the Timed "Up & Go" Test in People With Parkinson Disease," *Physical Therapy*, vol. 81, no. 2, pp. 810–818, Feb 2001, doi: 10.1093/ptj/81.2.810.

[5] B. R. Greene, B. Caulfield, D. Lamichhane, W. Bond, J. Svendsen, C. Zurski, and D. Pratt, "Longitudinal assessment of falls in patients with Parkinson's disease using inertial sensors and the Timed Up and Go test," *Journal of Rehabilitation and Assistive Technologies Engineering*, pp. 1–8, Jan 2018, doi: 10.1177/2055668317750811.

[6] R. N. Bellet, R. Francis, J. Jacob, K. Healy, H. Bartlett, L. Adams, and N. Morris, "Timed up and go tests in cardiac rehabilitation: reliability and comparison with the 6-minute walk test," *Journal of cardiopulmonary rehabilitation and prevention*, vol. 33, pp. 99–105, Mar 2012, doi: 10.1097/HCR.0b013e3182773fae.

[7] L. Filippin, F. Miraglia, J. Leite, R. Chakr, N. Oliveira, and D. Berwanger, "Identifying frailty syndrome with tug test in home-dwelling elderly," *Geriatrics, Gerontology and Aging*, vol. 11, pp. 80–87, May 2017, doi: 10.5327/Z2447-211520171700035.

[8] L. Kang, P. Han, J. Wang, Y. Ma, L. Jia, L. Fu, Y. Hairui, X. Chen, K. Niu, and Q. Guo, "Timed up and go test can predict recurrent falls: a longitudinal study of the community-dwelling elderly in china," *Clinical Interventions in Aging*, vol. Volume 12, p. 2009—2016, Nov 2017, doi: 10.2147/CIA.S138287.

[9] L. Rubenstein, "Falls in older people: Epidemiology, risk factors and strategies for prevention," *Age and ageing*, vol. 35 Suppl 2, Sep 2006, doi: 10.1093/ageing/afl084.

[10] E. Kramarow, L. H. Chen, H. Hedegaard, and M. Warner, "Deaths from unintentional injury among adults aged 65 and over: United states, 2000-2013," *NCHS data brief*, pp. 1–8, May 2015, pmid: 25973998.

[11] D. Schoene, C. Heller, Y. N. Aung, C. C. Sieber, W. Kemmler, and E. Freiberger, "A systematic review on the influence of fear of falling on quality of life in older people: Is there a role for falls?" *Clinical Interventions in Aging*, vol. Volume 14, pp. 701–719, Apr 2019, doi: 10.2147/CIA.S197857.

[12] S. Richardson, "The Timed "Up & Go": A Test of Basic Functional Mobility for Frail Elderly Persons," *Journal of the American Geriatrics Society*, vol. Volume 39(2), pp. 142–148, Feb 1991, doi: 10.1111/j.1532-5415.1991.tb01616.x.

[13] S. S. Ng and C. W. Hui-Chan, “The timed up & go test: Its reliability and association with lower-limb impairments and locomotor capacities in people with chronic stroke,” *Archives of Physical Medicine and Rehabilitation*, vol. 86, no. 8, pp. 1641–1647, Aug 2005, doi: 10.1016/j.apmr.2005.01.011.

[14] H. J. Van Hedel, M. Wirz, and V. Dietz, “Assessing walking ability in subjects with spinal cord injury: Validity and reliability of 3 walking tests,” *Archives of Physical Medicine and Rehabilitation*, vol. 86(2), pp. 190–6, 2005, doi: 10.1016/j.apmr.2004.02.010.

[15] G. M. Savva, O. A. Donoghue, F. Horgan, C. O’Regan, H. Cronin, and R. A. Kenny, “Using timed up-and-go to identify frail members of the older population,” *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, vol. 68, pp. 441–446, April 2013, doi: 10.1093/gerona/gls190.

[16] T. Herman, N. Giladi, and J. M. Hausdorff, “Properties of the ’Timed Up and Go’ test: More than meets the eye,” *Gerontology*, vol. 57(3), pp. 203–210, May 2011, doi: 10.1159/000314963.

[17] J. M. Hausdorff, G. Yogev, S. Springer, E. S. Simon, and N. Giladi, “Walking is more like catching than tapping: Gait in the elderly as a complex cognitive task,” *Experimental Brain Research*, vol. 164(4), pp. 541–8, Aug 2005, doi: 10.1007/s00221-005-2280-3.

[18] E. N. Picone, “The timed up and go test,” *American Journal of Nursing*, vol. 113(3), pp. 56–9, Mar 2013, doi: 10.1097/01.NAJ.0000427881.33418.cb.

[19] J. Ansai, A. Farche, P. Giusti Rossi, L. Andrade, T. Nakagawa, and A. Takahashi, “Performance of different timed up and go subtasks in frailty syndrome,” *Journal of Geriatric Physical Therapy*, vol. 42(4), pp. 287–293, Dec 2017, doi: 10.1519/JPT.0000000000000162.

[20] M. Osoba, A. Rao, S. Agrawal, and A. Lalwani, "Balance and gait in the elderly: A contemporary review: Balance and gait in the elderly," *Laryngoscope Investigative Otolaryngology*, vol. 4, pp. 143–153, Feb 2019, doi: 10.1002/lio2.252.

[21] J. Verghese, A. LeValley, C. Hall, M. Katz, A. Ambrose, and R. Lipton, "Epidemiology of gait disorders in community-residing older adults," *Journal of the American Geriatrics Society*, vol. 54(2), pp. 255–61, Feb 2006, doi: 10.1111/j.1532-5415.2005.00580.x.

[22] G. Scivoletto, F. Tamburella, L. Laurenza, C. Foti, J. Ditunno, and M. Molinari, "Validity and reliability of the 10-m walk test and the 6-min walk test in spinal cord injury patients," *Spinal cord*, vol. 49(6), pp. 736–40, Aug 2011, doi: 10.1038/sc.2010.180.

[23] A. Dunsky, A. Zeev, and Y. Netz, "Balance performance is task specific in older adults," *BioMed research international*, vol. 2017, pp. 1–7, Sept 2017, doi: 10.1155/2017/6987017.

[24] J. Silva and I. Sousa, "Instrumented timed up and go: Fall risk assessment based on inertial wearable sensors," *2016 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2016 - Proceedings*, vol. 2020, pp. 1–6, May 2016, doi: 10.1109/MeMeA.2016.7533778.

[25] K. Jahn, E. Freiberger, B. M. Eskofier, C. Bollheimer, and J. Klucken, "Balance and mobility in geriatric patients: Assessment and treatment of neurological aspects," pp. 316–323, Jul 2019, doi: 10.1007/s00391-019-01561-z.

[26] Statistics Canada, "The Cost of Injury in Canada," *Great West Life: Parachute Canada: Preventing Injuries. Saving Lives.*, 2015.

[27] C. Zampieri, A. Salarian, P. Carlson-Kuhta, J. G. Nutt, and F. B. Horak, "Assessing mobility at home in people with early Parkinson's disease using an

instrumented Timed Up and Go test," *Parkinsonism and Related Disorders*, vol. 17(4), pp. 277–280, May 2011, doi: 10.1016/j.parkreldis.2010.08.001.

[28] M. Montero-Odasso, J. Verghese, O. Beauchet, and J. M. Hausdorff, "Gait and cognition: a complementary approach to understanding brain function and the risk of falling," *Journal of the American Geriatrics Society*, vol. 60, no. 11, pp. 2127–2136, Nov 2012, doi: 10.1111/j.1532-5415.2012.04209.x.

[29] T. Doi, H. Shimada, H. Makizako, K. Tsutsumimoto, K. Uemura, Y. Anan, and T. Suzuki, "Cognitive function and gait speed under normal and dual-task walking among older adults with mild cognitive impairment," *BMC Neurology*, vol. 14, no. 1, p. 67, 2014, doi: 10.1186/1471-2377-14-67.

[30] W. Pirker and R. Katzenschlager, "Gait disorders in adults and the elderly : A clinical guide," *Wiener klinische Wochenschrift*, vol. 129, no. 3-4, pp. 81–95, Oct 2017, doi: 10.1007/s00508-016-1096-4.

[31] P. Gray and K. Hildebrand, "Fall risk factors in parkinson's disease," *Journal of Neuroscience Nursing*, vol. 32, no. 4, pp. 222–228, Aug 2000, doi : 10.1097/01376517-200008000-00006.

[32] A. Middleton and S. Fritz, "Assessment of gait, balance, and mobility in older adults: Considerations for clinicians," *Current Translational Geriatrics and Experimental Gerontology Reports*, vol. 2, pp. 205–214, Dec 2013, doi: 10.1007/s13670-013-0057-2.

[33] A. Schrag, M. Jahanshahi, N. MD, and N. Quinn, "How does parkinson's disease affect quality of life? a comparison with quality of life in the general population," *Movement Disorders*, vol. 15, pp. 1112–1118, Nov 2000, doi: 10.1002/1531-8257(200011)15:6¡1112::AID-MDS1008¿3.0.CO;2-A.

[34] R. Savica, A. M. V. Wennberg, C. Hagen, K. Edwards, R. O. Roberts, J. H. Hollman, D. S. Knopman, B. F. Boeve, M. M. Machulda, R. C. Petersen, and M. M. Mielke, "Comparison of gait parameters for predicting cognitive decline: The mayo clinic study of aging," *Journal of Alzheimer's disease : JAD*, vol. 55, no. 2, pp. 559–567, 2017, doi: 10.3233/JAD-160697.

[35] J. Verghese, G. Kuslansky, R. Holtzer, M. Katz, X. Xue, H. Buschke, and M. Pahor, "Walking While Talking: Effect of Task Prioritization in the Elderly," *Archives of Physical Medicine and Rehabilitation*, vol. 88(1), pp. 50–53, Jan 2007, doi: 10.1016/j.apmr.2006.10.007.

[36] G. Sprint, D. J. Cook, and D. L. Weeks, "Quantitative assessment of lower limb and cane movement with wearable inertial sensors," in *3rd IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2016*, Las Vegas, NV, USA, Feb 2016, pp. 418–421, doi: 10.1109/BHI.2016.7455923.

[37] N. Haji Ghassemi, J. Hannink, C. F. Martindale, H. Gaßner, M. Müller, J. Klucken, and B. M. Eskofier, "Segmentation of gait sequences in sensor-based movement analysis: A comparison of methods in parkinson's disease," *Sensors (Basel, Switzerland)*, vol. 18, no. 1, pp. 145–153, Nov 2018, doi: 10.3390/s18010145.

[38] A. Q. Abbasi and M. Abbasi, "Gait analysis using stride interval of left and right foot in single time series under normal and neurodegenerative diseased conditions," *Gait & Posture*, vol. 81(1), p. 1, Sep 2020, doi: 10.1016/j.gaitpost.2020.07.022.

[39] S.-U. Ko, K. Gunter, M. Costello, H. Aum, S. MacDonald, K. White, C. Snow, and W. Hayes, "Stride width discriminates gait of side-fallers compared to

other-directed fallers during overground walking," *Journal of aging and health*, vol. 19, 2007, doi: 10.1177/0898264307299308.

[40] N. Peel, S. Kuys, and K. Klein, "Gait speed as a measure in geriatric assessment in clinical settings: A systematic review," *The journals of gerontology. Series A, Biological sciences and medical sciences*, vol. 68(1), pp. 39–46, Jan 2012, doi: 10.1093/gerona/gls174.

[41] R. Chee, A. Murphy, M. Danoudis, N. Georgiou-Karistianis, and R. Iansek, "Gait freezing in parkinson's disease and the stride length sequence effect interaction," *Brain : a journal of neurology*, vol. 132, pp. 2151–60, June 2009, doi: 10.1093/brain/awp053.

[42] V. Agostini, G. Balestra, and M. Knaflitz, "Segmentation and classification of gait cycles," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 22, 2013, doi: 10.1109/TNSRE.2013.2291907.

[43] M. O. Derawi, P. Bours, and K. Holien, "Improved cycle detection for accelerometer based gait authentication," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, vol. 84, Oct 2010, pp. 312–317, doi: 10.1109/IIHMSP.2010.84.

[44] R. Libby, "A simple method for reliable footstep detection on embedded sensor platforms," *Sensors (Peterborough, NH)*, vol. 15(3), pp. 6419–6440, Jan 2008.

[45] S. M. M. De Rossi, S. Crea, M. Donati, P. Reberšek, D. Novak, N. Vitiello, T. Lenzi, J. Podobnik, M. Munih, and M. C. Carrozza, "Gait segmentation using bipedal foot pressure patterns," in *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2012, doi: 10.1109/BioRob.2012.6290278.

[46] V. Agostini, L. Gastaldi, V. Rosso, M. Knaflitz, and S. Tadano, "A wearable magneto-inertial system for gait analysis (h-gait): Validation on normal weight and overweight/obese young healthy adults," *Sensors*, vol. 17, no. 10, p. 2406, Oct 2017, doi: 10.3390/s17102406.

[47] S. Jiang, X. Wang, M. Kyrarini, and A. Graeser, "A robust algorithm for gait cycle segmentation," Kos, Greece, Oct 2017, pp. 31–35, doi: 10.23919/EUSIPCO.2017.8081163.

[48] A. Salarian, H. Russmann, F. Vingerhoets, C. Dehollain, Y. Blanc, P. Burkhard, and K. Aminian, "Gait assessment in parkinson's disease: toward an ambulatory system for long-term monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 8, pp. 1434 – 1443, July 2004, doi: 10.1109/TBME.2004.827933.

[49] S. R. Hundza, W. R. Hook, C. R. Harris, S. V. Mahajan, P. A. Leslie, C. A. Spani, L. G. Spalteholz, B. J. Birch, D. T. Commandeur, and N. J. Livingston, "Accurate and reliable gait cycle detection in parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 127–37, Jan 2014, doi: 10.1109/TNSRE.2013.2282080.

[50] M. E. Micó-Amigo, I. Kingma, E. Ainsworth, S. Walgaard, M. Niessen, R. Van Lummel, and J. Van Dieen, "A novel accelerometry-based algorithm for the detection of step durations over short episodes of gait in healthy elderly," *Journal of NeuroEngineering and Rehabilitation*, vol. 13, pp. 1–12, Apr 2016, doi: 10.1186/s12984-016-0145-6.

[51] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek, *Automatic Step Detection in the Accelerometer Signal*. Springer Berlin Heidelberg, June 2007, vol. 13, pp. 80–85, doi: 10.1007/978-3-540-70994-7-14.

[52] J. Barth, C. Oberndorfer, P. Kugler, D. Schuldhaus, J. Winkler, J. Klucken, and B. Eskofier, "Subsequence dynamic time warping as a method for robust step segmentation using gyroscope signals of daily life activities," vol. 2013, Osaka, Japan, July 2013, pp. 6744–6747, doi: 10.1109/EMBC.2013.6611104.

[53] P. De Leon and M. Martinez, "Unsupervised segmentation and labeling for smartphone acquired gait data," in *International Telemetering Conference Proceedings*, vol. 52, Jan 2016, http://hdl.handle.net/10150/624183.

[54] S. Misu, T. Asai, T. Doi, R. Sawa, Y. Ueda, S. Murata, T. Saito, T. Sugimoto, T. Isa, Y. Tsuboi, M. Yamada, and R. Ono, "Development and validation of Comprehensive Gait Assessment using InerTial Sensor score (C-GAITS score) derived from acceleration and angular velocity data at heel and lower trunk among community-dwelling older adults," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, pp. 1–10, May 2019, doi: 10.1186/s12984-019-0539-3.

[55] S. Potluri, A. Chandran, C. Diedrich, and L. Schega, "Machine learning based human gait segmentation with wearable sensor platform," vol. 2019, July 2019, pp. 588–594, doi: 10.1109/EMBC.2019.8857509.

[56] A. Prado, X. Cao, M. Robert, A. Gordon, and S. Agrawal, "Gait segmentation of data collected by instrumented shoes using a recurrent neural network classifier," *Physical Medicine and Rehabilitation Clinics of North America*, vol. 30, pp. 355–366, May 2019, doi: 10.1016/j.pmr.2018.12.007.

[57] M. Gadaleta, G. Cisotto, M. Rossi, R. z. u. Rehman, L. Rochester, and S. Din, "Deep learning techniques for improving digital gait segmentation," pp. 1834–1837, July 2019, doi: 10.1109/EMBC.2019.8856685.

[58] F. Tahavori, E. Stack, V. Agarwal, M. Burnett, A. Ashburn, S. A. Hoseinitabatabaei, and W. Harwin, "Physical activity recognition of elderly people and people with parkinson's (pwp) during standard mobility tests using wearable sensors," in *2017 International Smart Cities Conference (ISC2)*, Sept 2017, pp. 1–4, doi: 10.1109/ISC2.2017.8090858.

[59] D. Liciotti, G. Massi, E. Frontoni, A. Mancini, and P. Zingaretti, "Human activity analysis for in-home fall risk assessment," June 2015, pp. 284–289, doi: 10.1109/ICCW.2015.7247192.

[60] M. Babiker, O. O. Khalifa, K. K. Htike, A. Hassan, and M. Zaharadeen, "Automated daily human activity recognition for video surveillance using neural network," in *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, Putrajaya, Malaysia, November 2017, pp. 1–5, doi: 10.1109/ICSIMA.2017.8312024.

[61] P. Casale, O. Pujol, and P. Radeva, "Personalization and user verification in wearable systems using biometric walking patterns," *Personal and Ubiquitous Computing - PUC*, vol. 16, pp. 1617–4917, June 2012, doi: 10.1007/s00779-011-0415-z.

[62] J. Gupta, N. Singh, P. Dixit, V. Semwal, and S. Dubey, "Human activity recognition using gait pattern," p. 31–53, Jul 2013, doi: 10.4018/ijcvip.2013070103.

[63] R. Liu, S. A. Fazio, H. Zhang, A. A. Ramli, X. Liu, and J. Y. Adams, "Early mobility recognition for intensive care unit patients using accelerometers," Jun 2021.

[64] A. A. Ramli, H. Zhang, J. Hou, R. Liu, X. Liu, A. Nicorici, D. Aranki, C. Owens, P. Prasad, C. McDonald, and E. Henricson, "Gait characterization in duchenne

muscular dystrophy (dmd) using a single-sensor accelerometer: Classical machine learning and deep learning approaches," pp. 1–12, Jul 2021.

[65] M. Nutter, C. H. Crawford, and J. Ortiz, "Design of novel deep learning models for real-time human activity recognition with mobile phones," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, July 2018, pp. 1–8, doi: 10.1109/IJCNN.2018.8489319.

[66] M. Siddiqi, R. Ali, M. Rana, E.-K. Hong, E. Kim, and S. Lee, "Video-Based Human Activity Recognition Using Multilevel Wavelet Decomposition and Stepwise Linear Discriminant Analysis," *Sensors*, vol. 14, no. 4, pp. 6370–6392, March 2014, doi: 10.3390/s140406370.

[67] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik, "Human pose estimation with iterative error feedback," vol. V2, pp. 4733–4742, July 2015.

[68] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," pp. 1–14, 11 2016.

[69] T. Li, J. Chen, C. Hu, Y. Ma, Z. Wu, W. Wan, Y. Huang, F. Jia, C. Gong, S. Wan, and L. Li, "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.

[70] V. Bobić, M. Djuric-Jovicic, S. Radovanovic, N. Dragaevic, V. Kostić, and M. Popović, "Challenges of stride segmentation and their implementation for impaired gait," vol. 2018, Honolulu, HI, USA, July 2018, pp. 2284–2287, doi: 10.1109/EMBC.2018.8512836.

[71] G. Adusumilli, S. Lancia, V. Levasseur, V. Amblee, M. Orchard, J. Wagner, and R. Naismith, "Turning is an important marker of balance confidence and

walking limitation in persons with multiple sclerosis," *PLOS ONE*, vol. 13, p. e0198178, June 2018, doi: 10.1371/journal.pone.0198178.

[72] R. Almajid, R. Goel, C. Tucker, and E. Keshner, "Balance confidence and turning behavior as a measure of fall risk," *Gait & Posture*, vol. 80, pp. 1–6, May 2020, doi: 10.1016/j.gaitpost.2020.05.020.

[73] L. Khuna, P. Amatachaya, T. Sooknuan, T. Thaweewannakij, L. Mato, J. Seangsuwan, and S. Amatachaya, "Importance of independent sit-to-stand ability in ambulatory patients with spinal cord injury," *European journal of physical and rehabilitation medicine*, vol. 53, pp. 521–526, March 2017, doi: 10.23736/S1973-9087.17.04515-4.

[74] B.-M. Mun, Y.-S. Lee, T.-H. Kim, J.-H. Lee, S.-M. Sim, I.-M. Park, J. Park, and d.-k. Seo, "Study on the usefulness of sit to stand training in self-directed treatment of stroke patients," *Journal of physical therapy science*, vol. 26, pp. 483–485, April 2014, doi: 10.1589/jpts.26.483.

[75] M. Ghous, U. Rafi, M. Kanwal, and A. Malik, "Sit to stand is the precedent of balance and functional mobility in stroke," *journal of Riphah College of Rehabilitation Sciences*, vol. 2017; 5(2): 94-97, pp. 94–97, Oct 2017.

[76] A. Salarian, F. Horak, C. Zampieri, P. Carlson-Kuhta, J. Nutt, and K. Aminian, "itug, a sensitive and reliable measure of mobility," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 3, pp. 303–310, April 2010, doi: 10.1109/TNSRE.2010.2047606.

[77] H. Manaf, M. Justine, and M. Omar, "Functional balance and motor impairment correlations with gait parameters during timed up and go test across three attentional loading conditions in stroke survivors," *Stroke research and treatment*, vol. 2014, pp. 1–10, March 2014, doi: 10.1155/2014/439304.

[78] A. Alghadir, S. Anwer, and J.-M. Brismée, “The reliability and minimal detectable change of timed up and go test in individuals with grade 1 – 3 knee osteoarthritis,” *BMC Musculoskeletal Disorders*, vol. 16, no. 1, p. 174, Dec 2015, doi: 10.1186/s12891-015-0637-8.

[79] A. Wright, C. Cook, G. D. Baxter, J. Dockerty, and J. H. Abbott, “A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis,” *The Journal of orthopaedic and sports physical therapy*, vol. 41, pp. 319–27, Feb 2011, doi: 10.2519/jospt.2011.3515.

[80] W. Vieira, T. Ostolin, M. Ferreira, E. Sperandio, and V. Dourado, “Test timed up and go and its correlation with age and functional exercise capacity in asymptomatic women,” *Fisioterapia em Movimento*, vol. 30, pp. 463–471, Sept 2017, doi: 10.1590/1980-5918.030.003.ao04.

[81] A. Ziegl, D. Hayn, P. Kastner, K. Loeffler, L. Weidinger, B. Brix, N. Goswami, and G. Schreier, “Quantitative falls risk assessment in elderly people: results from a clinical study with distance based timed up-and-go test recordings,” *Physiological Measurement*, vol. 41, pp. 1–26, Oct 2020, doi: 10.1088/1361-6579/abc352.

[82] J. Ansai, L. Andrade, T. Nakagawa, F. Vale, M. J. Caetano, S. Lord, and J. Rebelatto, “Cognitive correlates of timed up and go subtasks in older people with preserved cognition, mild cognitive impairment, and alzheimer’s disease,” *American journal of physical medicine & rehabilitation*, vol. 96, pp. 1–6, Feb 2017, doi: 10.1097/PHM.0000000000000722.

[83] V. De Leon Gomez, C. Barone, Y. Aoustin, and C. Chevallereau, “Study of the Walking Efficiency of a Human with a Cane,” in *CISM International Centre*

for Mechanical Sciences, Courses and Lectures, January 2019, pp. 370–379, doi: 10.1007/978-3-319-78963-7-47.

[84] N. M. Gell, R. B. Wallace, A. Z. Lacroix, T. M. Mroz, and K. V. Patel, "Mobility device use in older adults and incidence of falls and worry about falling: Findings from the 2011-2012 national health and aging trends study," *Journal of the American Geriatrics Society*, pp. 853–859, May 2015, doi: 10.1111/jgs.13393.

[85] D. Vollmerdahlke and M. Ory, "Emerging issues of intelligent assistive technology use among people with dementia and their caregivers: A u.s. perspective," *Frontiers in Public Health*, vol. 8, pp. 1–7, May 2020, doi: 10.3389/fpubh.2020.00191.

[86] I. Y. Chung, S. Kim, and K. H. Rhee, "The smart cane utilizing a smart phone for the visually impaired person," in *2014 IEEE 3rd Global Conference on Consumer Electronics, GCCE 2014*, Oct 2014, pp. 106–107, doi: 10.1109/GCCE.2014.7031333.

[87] L. Matalenas, T. Hu, V. Veeramachaneni, and J. Muth, "A cane for more than walking: The design of a smart cane for physical therapy," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, pp. 1063–1067, 09 2016, doi: 10.1177/1541931213601246.

[88] S. Gill, N. Seth, and E. Scheme, "A multi-sensor cane can detect changes in gait caused by simulated gait abnormalities and walking terrains," *Sensors (Basel, Switzerland)*, vol. 20, p. 631, 01 2020.

[89] M. Lan, A. Nahapetian, A. Vahdatpour, L. Au, W. Kaiser, and M. Sarrafzadeh, "Smartfall: An automatic fall detection system based on subsequence matching for the smartcane," *BODYNETS 2009 - 4th Interna-*

tional ICST Conference on Body Area Networks, pp. 1–8, 05 2009, doi: 10.4108/ICST.BODYNETS2009.5873.

[90] J. Wade, M. Beccani, A. Myszka, E. Bekele, P. Valdastri, P. Flemming, M. De Riesthal, T. Withrow, and N. Sarkar, "Design and implementation of an instrumented cane for gait recognition," in *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 2015-June, no. June, June 2015, pp. 5904–5909, doi: 10.1109/ICRA.2015.7140026.

[91] 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.

[92] D. Berrada, M. Romero, G. Abowd, M. Blount, and J. Davis, "Automatic administration of the get up and go test," Jun 2007, pp. 73–75, doi: 10.1145/1248054.1248075.

[93] Z. Skrba, B. O'Mullane, B. R. Greene, C. N. Scanaill, C. W. Fan, A. Quigley, and P. Nixon, "Objective real-time assessment of walking and turning in elderly adults," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, September 2009, pp. 807–810, doi: 10.1109/IEMBS.2009.5333934.

[94] W. Durfee, L. Savard, and S. Weinstein, "Technical feasibility of teleassessments for rehabilitation," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 15, pp. 23–29, March 2007, doi: 10.1109/TNSRE.2007.891400.

[95] N. Kitsunezaki, E. Adachi, T. Masuda, and J.-i. Mizusawa, "Kinect applications for the physical rehabilitation," in *2013 IEEE International Symposium*

on Medical Measurements and Applications (MeMeA), Gatineau, QC, Canada, May 2013, pp. 294–299, doi: 10.1109/MeMeA.2013.6549755.

[96] F. Wang, M. Skubic, C. Abbott, and J. M. Keller, "Quantitative analysis of 180 degree turns for fall risk assessment using video sensors," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, MA, USA, Dec 2011, pp. 7606–7609, doi: 10.1109/IEMBS.2011.6092005.

[97] O. Lohmann, T. Luhmann, and A. Hein, "Skeleton timed up and go," in *2012 IEEE International Conference on Bioinformatics and Biomedicine*, Philadelphia, PA, USA, Dec 2012, pp. 1–5, doi: 10.1109/BIBM.2012.6392610.

[98] A. Dubois, T. Bihl, and J. P. Bresciani, "Automating the timed up and go test using a depth camera," *Sensors (Switzerland)*, vol. 18, p. 14, Dec 2018, doi: 10.3390/s18010014.

[99] M. R. Narayanan, S. R. Lord, M. M. Budge, B. G. Celler, and N. H. Lovell, "Falls management: Detection and prevention, using a waist-mounted triaxial accelerometer," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, France, Oct 2007, pp. 4037–4040, doi: 10.1109/IEMBS.2007.4353219.

[100] 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, July 2008, doi: 10.1109/MEMB.2008.919494.

[101] A. Weiss, T. Herman, M. Plotnik, M. Brozgol, N. Giladi, and J. M. Hausdorff, "An instrumented timed up and go: The added value of an accelerometer for

identifying fall risk in idiopathic fallers," *Physiological Measurement*, vol. 32, no. 12, pp. 2003–2018, Nov 2011, doi: 10.1088/0967-3334/32/12/009.

[102] A. Weiss, A. Mirelman, A. S. Buchman, D. A. Bennett, and J. M. Hausdorff, "Using a Body-Fixed Sensor to Identify Subclinical Gait Difficulties in Older Adults with IADL Disability: Maximizing the Output of the Timed Up and Go," *PLoS ONE*, vol. 8, no. 7, pp. 1–8, July 2013, doi: 10.1371/journal.pone.0068885.

[103] B. Najafi, K. Aminian, F. Loew, Y. Blanc, and P. Robert, "Measurement of stand-sit and sit-stand transitions using a miniature gyroscope and its application in fall risk evaluation in the elderly," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 8, pp. 843 – 851, August 2002, doi: 10.1109/TBME.2002.800763.

[104] A. Godfrey, A. Bourke, G. ÓLaighin, P. Ven, and J. Nelson, "Activity classification using a single chest mounted tri-axial accelerometer," *Medical engineering & physics*, vol. 33, no. 9, pp. 1127–35, May 2011, doi: 10.1016/j.medengphy.2011.05.002.

[105] 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," in *2019 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2019 - Proceedings*, Chicago, IL, USA, Sept 2019, pp. 1–4, doi: 10.1109/BHI.2019.8834646.

[106] H. P. Nguyen, F. Ayachi, C. Lavigne-Pelletier, M. Blamoutier, F. Rahimi, P. Boissy, M. Jog, and C. Duval, "Auto detection and segmentation of physical activities during a Timed-Up-and-Go (TUG) task in healthy older adults us-

ing multiple inertial sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, pp. 1–12, April 2015, doi: 10.1186/s12984-015-0026-4.

[107] T. Frenken, B. Vester, M. Brell, and A. Hein, "ATUG: Fully-automated timed up and go assessment using ambient sensor technologies," in *2011 5th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2011*, Dublin, Ireland, April 2011, pp. 55–62, doi: 10.4108/icst.pervasivehealth.2011.245985.

[108] M. Milosevic, E. Jovanov, and A. Milenković, "Quantifying timed-up-and-go test: A smartphone implementation," in *2013 IEEE International Conference on Body Sensor Networks, BSN 2013*, Cambridge, MA, USA, August 2013, pp. 1–6, doi: 10.1109/BSN.2013.6575478.

[109] A. F. R. Kleiner, I. Pacifici, A. Vagnini, F. Camerota, C. Celletti, F. Stocchi, M. F. De Pandis, and M. Galli, "Timed Up and Go evaluation with wearable devices: Validation in Parkinson's disease," *Journal of Bodywork and Movement Therapies*, vol. 22 2, pp. 390–395, July 2018, doi: 10.1016/j.jbmt.2017.07.006.

[110] M. R. Adame, A. Al-Jawad, M. Romanovas, M. A. Hobert, W. Maetzler, K. Möller, and Y. Manoli, "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, p. 1071–1074, August 2012, doi: doi:10.1515/bmt-2012-4426.

[111] L. Palmerini, S. Mellone, L. Rocchi, and L. Chiari, "Dimensionality reduction for the quantitative evaluation of a smartphone-based timed up and go test," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2011, Aug 2011, pp. 7179–82, doi: 10.1109/IEMBS.2011.6091814.

[112] 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, July 2011, pp. 689–692, doi: 10.1109/ICASSP.2011.5946497.

[113] S. Reinfelder, R. Hauer, J. Barth, J. Klucken, and B. M. Eskofier, "Timed Up-and-Go phase segmentation in Parkinson's disease patients using unobtrusive inertial sensors," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Milan, Italy, Aug 2015, pp. 5171–5174, doi: 10.1109/EMBC.2015.7319556.

[114] 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, p. 6302, Sept 2020.

[115] M. Ko, W. Geoff, S. Venkatesh, and M. Kumar, "Using dynamic time warping for online temporal fusion in multisensor systems," *Information Fusion*, vol. 9, pp. 370–388, July 2008, doi: 10.1016/j.inffus.2006.08.002.

[116] K. Tumilaar, Y. Langi, and A. Rindengan, "Hidden markov model," *d'CARTESIAN*, vol. 4, p. 86, Feb 2015, doi: 10.35799/dc.4.1.2015.8104.

[117] L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989, doi: 10.1109/5.18626.

[118] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, April 2015, p. 268, doi: 10.1007/978-1-4302-5990-9-5.

[119] P. Roy, P. Kumar, and B.-G. Kim, "An efficient sign language recognition (slr) system using camshift tracker and hidden markov model (hmm)," *SN Computer Science*, vol. 2, Feb 2021, doi: 10.1007/s42979-021-00485-z.

[120] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," in *Inequalities III: Proceedings of the Third Symposium on Inequalities*, O. Shisha, Ed. University of California, Los Angeles: Academic Press, 1972.

[121] Z. Hatala and V. Puturuhu, "Viterbi extraction tutorial with hidden markov toolkit," Aug 2019.

[122] Z. Zhao, R. Anand, and M. Wang, "Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform," in *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Washington, DC, USA, Jan 2019, pp. 442–452, doi: 10.1109/DSAA.2019.00059.

[123] C. Cooney, A. Korik, R. Folli, and D. Coyle, "Classification of imagined spoken word-pairs using convolutional neural networks," Jone 2019, doi: 10.3217/978-3-85125-682-6-62.

[124] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," April 2019.

[125] H. Tan, N. Aung, J. Tian, M. Chua, and Y. Ou Yang, "Time series classification using a modified lstm approach from accelerometer-based data: A comparative study for gait cycle detection," *Gait & Posture*, vol. 74, pp. 128–134, Oct 2019, doi: 10.1016/j.gaitpost.2019.09.007.

[126] M. Milenkoski, K. Trivodaliev, S. Kalajdziski, M. Jovanov, and B. Risteska Stojkoska, "Real time human activity recognition on smartphones using lstm networks," 2018, doi: 10.23919/MIPRO.2018.8400205.

[127] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A modified lstm model for continuous sign language recognition using leap motion," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7056–7063, Aug 2019, doi :doi: 10.1109/JSEN.2019.2909837.

[128] P. Kumar and E. Scheme, "A deep spatio-temporal model for eeg-based imagined speech recognition," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, doi: 10.1109/ICASSP39728.2021.9413989.

[129] K. Cho, B. van Merriënboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *EMNLP 2014*, June 2014, doi: 10.3115/v1/D14-1179.

[130] J. Xu, W. Luo, and Y. Huang, "Dadu river runoff forecasting via seq2seq," 2019, pp. 494–498, doi: 10.1145/3349341.3349457.

[131] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," April 2018, url: abs/1803.01271.

[132] S.-P. Lee, J. Dufek, R. Hickman, and S. Schuerman, "Influence of procedural factors on the reliability and performance of the timed up-and-go test in older adults," *International Journal of Gerontology*, vol. 10, no. 1, pp. 37–42, March 2016, doi: https://doi.org/10.1016/j.ijge.2015.10.003.

[133] 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), Athens, Greece (Virtual), Aug 2021, pp. 1–5, doi: 10.1109/BHI50953.2021.9508595.

Vita

Candidate's full name: Ameya Sharad Valsangkar

University attended (with dates and degrees obtained):

1. 2015 - Bachelor's of Engineering (Electronics and Telecommunication), University of Mumbai.

Publications:

1. Ameya Valsangkar, Pradeep Kumar, Erik Scheme, "Automated Segmentation of a Timed Up and Go Test Using an Instrumented Cane", 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), August 2021, DOI: 10.1109/BHI50953.2021.9508595 [133]