

Article

AI-Driven Adaptive Segmentation of Timed Up and Go Test Phases Using a Smartphone

Muntazir Rashid ¹, Arshad Sher ^{2,*}, Federico Villagra Povina ³ and Otar Akanyeti ⁴¹ Independent Researcher, Weston Super Mare BS24 9JL, UK; muntazirrashid50@gmail.com² Department of Computer Science, Nottingham Trent University, Nottinghamshire NG11 8NS, UK³ Department of Life Sciences, Aberystwyth University, Ceredigion SY23 3FL, UK⁴ Department of Computer Science, Aberystwyth University, Ceredigion SY23 3DB, UK

* Correspondence: arshad.sher@ntu.ac.uk

Abstract

The Timed Up and Go (TUG) test is a widely used clinical tool for assessing mobility and fall risk in older adults and individuals with neurological or musculoskeletal conditions. While it provides a quick measure of functional independence, traditional stopwatch-based timing offers only a single completion time and fails to reveal which movement phases contribute to impairment. This study presents a smartphone-based system that automatically segments the TUG test into distinct phases, delivering objective and low-cost biomarkers of lower-limb performance. This approach enables clinicians to identify phase-specific impairments in populations such as individuals with Parkinson's disease, and older adults, supporting precise diagnosis, personalized rehabilitation, and continuous monitoring of mobility decline and neuroplastic recovery. Our method combines adaptive preprocessing of accelerometer and gyroscope signals with supervised learning models (Random Forest, Support Vector Machine (SVM), and XGBoost) using statistical features to achieve continuous phase detection and maintain robustness against slow or irregular gait, accommodating individual variability. A threshold-based turn detection strategy captures both sharp and gradual rotations. Validation against video ground truth using group K-fold cross-validation demonstrated strong and consistent performance: start and end points were detected in 100% of trials. The mean absolute error for total time was 0.42 s (95% CI: 0.36–0.48 s). The average error across phases (stand, walk, turn) was less than 0.35 s, and macro F_1 scores exceeded 0.85 for all models, with the SVM achieving the highest score of 0.882. Combining accelerometer and gyroscope features improved macro F_1 by up to 12%. Statistical tests (McNemar, Bowker) confirmed significant differences between models, and calibration metrics indicated reliable probabilistic outputs (ROC-AUC > 0.96, Brier score < 0.08). These findings show that a single smartphone can deliver accurate, interpretable, and phase-aware TUG analysis without complex multi-sensor setups, enabling practical and scalable mobility assessment for clinical use.

Keywords: Timed Up and Go; smartphone sensors; supervised learning; accelerometer; gyroscope; turning; walking



Academic Editor: Gianluca Traversi

Received: 20 October 2025

Revised: 21 November 2025

Accepted: 24 November 2025

Published: 26 November 2025

Citation: Rashid, M.; Sher, A.; Povina, F.V.; Akanyeti, O. AI-Driven Adaptive Segmentation of Timed Up and Go Test Phases Using a Smartphone. *Electronics* **2025**, *14*, 4650. <https://doi.org/10.3390/electronics14234650>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Mobility decline is a common issue among older adults and is closely linked to fall risk, loss of independence, and reduced quality of life [1–4]. Clinical functional assessments often rely on subjective evaluations and a stopwatch [4,5]. In the Timed Up and Go (TUG)

test, clinicians typically measure only the total completion time from the moment the participant starts with standing from a chair, walks 3 m, takes a turn, walks back to the chair, and sits back down [6]. The test completion time is used to estimate mobility quality and independence [7–9]. Another widely used tool, the Modified Rankin Scale, assigns a score from 0 (no mobility issues) to 6 (severe impairment) [10].

Manual timing is prone to error and provides limited insight, whereas smartphones equipped with inertial sensors can capture detailed motion data [11–13]. Two patients may have the same overall TUG time but very different movement patterns: one may struggle to stand up, while another may have difficulty turning. Knowing which phase takes longer helps clinicians identify the source of impairment and design targeted interventions, such as strength training for sit-to-stand or balance exercises for turning. Phase-level information can also track subtle changes over time, supporting early detection of deterioration and personalized rehabilitation planning [8,9].

Recent advances in smartphone technologies highlight the importance of automating traditional approaches [11,14–20]. This enables automated detection of test start and end points and, more importantly, segmentation of sub-phases, which is essential for understanding where mobility challenges occur. For example, prolonged sit-to-stand may indicate lower limb weakness, while extended turning time may suggest balance or vestibular issues. Such insights go beyond total time and provide actionable information for clinical decision-making. Existing studies have validated instrumented TUG using wearable sensors, but most work has focused on controlled environments [21–24]. Real-world monitoring requires adaptability to variations in age, sex, test settings, and individual movement styles, making phase-level analysis even more critical for personalized care [5,10,25,26]. Continuous monitoring could also reveal changes in phase timings over time, reflecting fatigue, daily routines, or underlying health conditions [10].

The motivation for this work is to introduce a smartphone-based system that automatically segments the Timed Up and Go test into distinct phases, providing objective, low-cost biomarkers of lower-limb performance. Unlike traditional stopwatch-based timing, our approach enables clinicians to identify phase-specific impairments in Parkinson’s disease and older adults, supporting precise diagnosis, personalized rehabilitation, and continuous monitoring of mobility decline and neuroplastic recovery. In this study, we propose a smartphone-based system with the following contributions:

- Adaptive preprocessing: Thresholds derived from median and median absolute deviation (MAD) per trial adjust sensitivity to individual movement amplitude and enforce a no-gaps policy, preventing fragmentation during slow or irregular walking. This adaptability ensures reliable segmentation across diverse gait patterns, which is critical for clinical use.
- Sensor fusion: Accelerometer and gyroscope signals are combined to capture both linear and rotational dynamics, improving detection of walking and turning phases. This enables an accurate phase-level analysis, which helps clinicians identify whether difficulties arise from straight walking or turning.
- Adaptive turn detection: A peak-based path and an angle-area path work together to identify both sharp and gradual rotations, addressing a key limitation in prior single-device approaches. Accurate turn detection is clinically important because turning deficits are strongly associated with fall risk.
- Statistical features and classical models: Features summarizing magnitude, variability, and energy are extracted from short windows and classified using Random Forest, Support Vector Machine (SVM), and XGBoost, ensuring interpretability and efficiency. This design supports practical deployment on consumer devices without sacrificing accuracy.

Our work is validated against video-based ground truth, reporting start and end detection in 100% of trials, a mean absolute error of 0.42 s for total time, an average phase-level mean absolute error below 0.35 s, and macro F1 scores above 0.85 across all models. These results demonstrate that a single consumer device can deliver accurate, phase-aware TUG analysis without complex multi-sensor setups, advancing automated mobility assessment toward practical deployment.

The remainder of this paper is organized as follows: Section 2 reviews related work on sensor-based TUG automation. Section 3 provides the details on the proposed methodology. Sections 4 and 5 present results and discussion. Finally, Section 6 concludes the study with future directions.

2. Related Work

Existing studies on automating the TUG test have shown that it is feasible to use both wearable sensors and smartphones for automating the TUG test. Research using wearable sensors often places multiple devices on different body parts to capture detailed motion data [27–30], while smartphone-based approaches aim for a simpler and more practical solution for real-world use. Both strategies have demonstrated strong potential but need more exploration to address gaps, particularly adaptive methods that can adjust to individual differences in mobility. This adaptability is essential for handling variability in real-world conditions and for providing accurate, quantitative measures of participant performance.

One of the early works conducted by Salarian et al. [21] proposed an algorithm to automatically analyze the subcomponents of the instrumented TUG (iTUG) test. This study enrolled 12 subjects with early-stage idiopathic Parkinson’s disease (60.4 ± 8.5 years) and 12 age-matched controls (60.2 ± 8.2 years), who performed a modified iTUG over a 7-m distance. Completion time was measured using a stopwatch, and video recordings were used for verification. Inertial sensors were placed on multiple body locations, including the forearms, shanks, thighs, lower back, and sternum, to capture movement data. A subgroup of nine participants was included to assess test-retest reliability. The system identified four subcomponents: sit-to-stand, steady-state walking, turning, and stand-to-sit. This study reported accuracy above 0.75 for spatial and above 0.90 for temporal gait parameters, but the setup was complex and not practical for real-world settings.

In order to make the TUG test automation setup easy, Ishikawa et al. [31] analysed gait during the TUG test using a smartphone application to evaluate six components: stand, walk, turn1, walk, turn2, and sit. The study involved 87 older adults (stroke, cardiac disorder, hip fracture) and 32 participants with idiopathic normal pressure hydrocephalus (iNPH). An iPhone was placed on the abdomen to record iTUG at the fastest possible walking speed. The system showed excellent inter-class correlation (ICC) with manual measurements (ICC = 0.93 for manual TUG and ICC = 0.94 for iTUG). Completion times were 12.5 s for active participants and 19.4 s for iNPH in iTUG, compared to 10.9 s and 18.1 s in manual TUG. Validation against video annotations achieved 87.4% accuracy for active participants and 81.2% for iNPH.

Clavijo-Buendía et al. [32] developed a free mobile application for spatio-temporal gait analysis in individuals with Parkinson’s disease. The app was mounted on the anterior thigh of the affected side and tested on 30 participants (71.7 ± 5.1 years) during a 10-m walk. Parameters were measured using both a stopwatch and the app (RUNZI) within two hours of medication intake. The study also included iTUG, Tinetti Scale, and Berg Balance Scale for construct validity. The primary objective was to compare gait parameters between the 10-m walk and TUG, but phase-level segmentation for TUG was not addressed. More recent work by Matey-Sanz et al. [15] combined smartphones and smartwatches to automate TUG, while Böttinger et al. [16] explored self-administered TUG using a smartphone. These

studies improve accessibility but still focus mainly on total time and do not fully solve phase segmentation or turn detection.

Mellone et al. [30] examined the validity of a smartphone-based iTUG in 49 participants without specific inclusion or exclusion criteria. The smartphone was compared against the McRoberts Dynaport Hybrid, a research-grade device. Both devices were attached to the lower back and recorded data simultaneously. The study identified heel strikes and computed step time, sit-to-stand duration, and total completion time. Intra-rater reliability was 69.75, and inter-rater reliability was 94.25.

Turki S. Abualait et al. [33] investigated the effect of assistive walking devices on modified iTUG performance in 20 healthy adults (22.8 ± 4.42 years). The study analyzed spatio-temporal and functional mobility parameters, reporting significant differences in stride velocity, stride length, and cadence ($p = 0.001$) when walking with and without aid. Sit-to-stand time was slower compared to walking with a walker ($p = 0.004$) or a cane ($p = 0.004$). No significant difference was observed between the cane and the four-wheeled walker ($p = 0.94$). The findings indicate that assistive devices alter gait by increasing stride time and reducing cadence, but the study did not address automation or remote monitoring. A summary of the reviewed related studies is provided in Table 1.

Table 1. Summary of related work studies closely related to our work on TUG automation.

Study	Device and Placement	What They Achieved	Limitations
Salarian et al. [21]	Multiple IMUs on limbs and trunk	Accurate phase detection and gait metrics	Complex setup, not practical for home use
Ortega-Bastidas et al. [14]	Single IMU on lower back	Segmentation for walking phases	Weak turn detection; no adaptive thresholds
Matey-Sanz et al. [15]	Smartphone + smartwatch	Automated TUG with better usability	Requires multiple devices
Ishikawa et al. [31]	Smartphone on abdomen	Six-phase segmentation; ICC ≈ 0.94	Limited adaptability to variable gait
Mellone et al. [30]	Smartphone on lower back + reference device	Valid total time and sit-to-stand detection	Minimal phase-level detail

Existing studies are either focusing on multiple sensors or assume controlled conditions [14–16,30,32]. Few combine simplicity with adaptability. Our work addresses this gap by using a single smartphone at the lower back, adaptive preprocessing to handle quiet starts and curvilinear turns, sensor fusion for better phase detection, and supervised learning for robust classification. We validate the approach against video ground truth using group K-fold cross-validation and support feasibility for practical deployment.

3. Methodology

A custom Android application was developed to record motion data using the smartphone's built-in inertial sensors. Before each trial, the experimenter entered participant ID and demographics (age, height, and weight) and selected activity type, trial number, and recording duration from drop-down menus. Data capture was started manually. At the end of each session, sensor data and metadata were combined into a single JSON file

and uploaded to a secure server at Aberystwyth University. If no internet connection was available, the file was stored locally and uploaded automatically once connectivity was restored.

3.1. Participants

A total of 27 volunteers (13 males, 14 females; age: 71.5 ± 12.07 years; weight: 67.8 ± 14.36 kg; height: 165.5 ± 10.2 cm) were recruited. Participants were classified as frail based on slow walking speed ($<0.8 \text{ m s}^{-1}$) and prolonged TUG completion time ($>10 \text{ s}$) [6]. All participants were able to walk independently. Five participants had Parkinson's disease, while the remaining participants had no additional neurological or chronic physical conditions that could affect gait during the TUG test. Sample sizes were determined based on recruitment feasibility and the study's focus on older adults ($n = 22$), for whom TUG test performance is most clinically relevant. A Parkinson's disease subgroup ($n = 5$) was included to assess the feasibility and adaptability of the proposed algorithm within a clinical cohort. These allocations reflect practical constraints and the exploratory nature of the study.

While our smartphone-based system has been validated on diverse subject datasets, comprising older adults and patients with Parkinson's disease, the current implementation does not explicitly differentiate between these groups during segmentation. However, the adaptive preprocessing and phase-level analysis are designed to generalize across heterogeneous gait patterns, making the approach suitable for future group-specific modeling.

All participants provided informed consent before participation. The study was approved by Aberystwyth University and NHS Ethical Committees and conducted in accordance with the Declaration of Helsinki.

3.2. Data Collection

TUG Protocol: Participants performed the standard TUG test, which consists of rising from a chair, walking three meters, turning 180 degrees, walking back, and sitting down while turning 180 degrees, as illustrated in Figure 1.

TUG tests were recorded using a Google Pixel 2 smartphone at an average sampling rate of 405 Hz. The device was secured at the lower back near the L3 vertebra using a fixation belt to minimize motion artifacts. While the device size may influence participant behavior due to its placement, this location was chosen because it approximates the body's center of mass and is widely recommended in IMU-based gait analysis. It provides stable axes for acceleration and angular velocity, reducing orientation variability compared to pocket placement. We acknowledge that this choice does not fully eliminate the possibility of altered movement due to device awareness. However, prior validation studies confirm that lumbar-mounted smartphones achieve excellent reliability and validity for spatiotemporal gait parameters ($\text{ICC} \geq 0.90$, $r \geq 0.89$) across walking speeds [29]. Although front-pocket and shoulder bag placement remain a promising option for real-world deployment, they introduce variability in sensor orientation and body coupling, factors not addressed in the present study. As part of future work, we plan to collect pocket-mounted trials and develop placement-aware models to evaluate performance equivalence with the L3 configuration. Furthermore, existing evidence indicates that a smartphone positioned at L3 does not materially alter gait, and trunk kinematics measured at this position are accurate within $5\text{--}10^\circ$ of motion capture systems [27]. The smartphone's orientation was standardized across all trials: the accelerometer's z -axis aligned with the anterior–posterior direction, the y -axis with the medial–lateral direction, and the x -axis represented vertical acceleration. Data collection began approximately five seconds before and ended five seconds after each trial, and all uploaded data were visually inspected for quality control.

Ground Truth Recording: Each trial was recorded using a GoPro Hero 6 (240 fps, 1920×1080 resolution) mounted on a tripod at a height of 56 cm to capture the entire walking path. Videos were reviewed frame by frame to ensure precise identification of transitions, including the start and end of sit-to-stand, walking, turning, and stand-to-sit phases. These annotations provided ground truth for computing phase durations and total TUG time. The manually extracted timings served as the reference for validating algorithm outputs. Agreement between algorithm-derived and video-derived timings was quantified using mean absolute error (MAE).

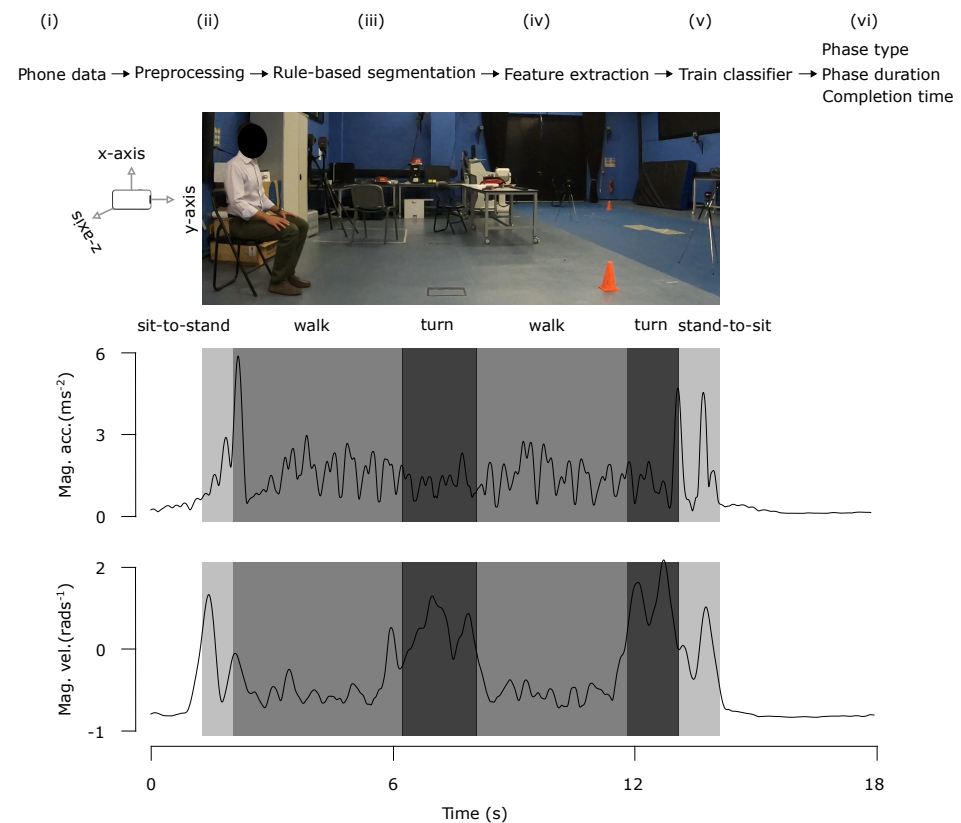


Figure 1. Overview of the smartphone-based pipeline for TUG segmentation. The top row illustrates the processing stages: (i) data acquisition from the smartphone, (ii) preprocessing, (iii) rule-based segmentation, (iv) feature extraction, (v) classifier training, and (vi) output generation (TUG phase type, phase duration, and total completion time). Below, an image of the actual TUG test demonstrates the setup, including the chair and the 3-m turning cone. Underneath, gyroscope and accelerometer traces are shown with shaded regions corresponding to the five core phases: (i) sit-to-stand, (ii) walk forward, (iii) turn, (iv) walk back, and (v) stand-to-sit. Accelerometer signals capture linear motion during walking and transitions, while gyroscope signals highlight rotational motion during turns.

3.3. Data Processing Pipeline

The pipeline consists of processing, rule-based segmentation, feature extraction, classifier training and testing, and phase type, duration and total time estimation. The framework is shown in Figure 1.

3.3.1. Processing

The raw accelerometer and gyroscope signals are filtered to reduce noise and make them suitable for phase detection [21,22,31]. A low-pass filter is applied to each channel (cut-off frequency 6 Hz for acceleration and 6 Hz for angular velocity) to remove high-frequency components while preserving gait dynamics [8]. After filtering, we compute magnitude envelopes and remove the gravity component using a zero-phase 2nd-order

Butterworth lowpass baseline filter at 0.25 Hz. This isolates dynamic acceleration, which is essential for detecting transitions such as sit-to-stand and turning.

To capture local energy patterns that characterize walking and turning, we calculate root mean square (RMS) envelopes using short moving averages: 150 ms for acceleration and 200 ms for gyroscope signals. The window sizes differ because acceleration changes more rapidly, especially during heel strikes and sit-to-stand transitions, so a shorter window (150 ms) preserves temporal sensitivity without over-smoothing. In contrast, angular velocity varies more smoothly during rotation, so a slightly longer window (200 ms) provides stability for turn detection and reduces false peaks. These RMS envelopes provide a smoothed representation of motion intensity, making it easier to distinguish between active and still phases of TUG. After that, we resample all envelopes to a common time base, using the gyroscope as the reference, to ensure alignment across both signals. This design balances temporal sensitivity for accelerometer data and stability for gyroscope data.

Finally, we compute adaptive thresholds based on the median and median absolute deviation (MAD) of each trial. This choice is motivated by robustness: median and MAD are less sensitive to outliers and skewed distributions than mean and standard deviation, which is critical because gait signals often include irregular peaks from slow walking or brief pauses.

Adaptive thresholds were computed to detect significant motion events during the TUG test. For the accelerometer signal, the threshold $th_{acc}(t)$ was based on the envelope of the acceleration magnitude (Acc_{env}) and defined as:

$$th_{acc}(t) = \begin{cases} \text{median}(Acc_{env}) + 2.0 \times \text{MAD}, & t \leq t_0 + 1.5 \text{ s}, \\ \text{median}(Acc_{env}) + 2.25 \times \text{MAD}, & \text{otherwise.} \end{cases}$$

Here, $\text{median}(Acc_{env})$ is the median of the accelerometer envelope, and MAD is the Median Absolute Deviation. A lower multiplier (2.0) is applied during the first 1.5 s after the start time t_0 to capture early movements, while a slightly higher multiplier (2.25) is used afterward to reduce false positives. For the gyroscope signal, the threshold th_{gyr} was computed from the root mean square of the gyroscope signal (Gyr_{rms}) as:

$$th_{gyr} = \text{median}(Gyr_{rms}) + 2.0 \times \text{MAD}.$$

These adaptive thresholds leverage robust statistics (median and MAD) to accommodate individual variability and noise, ensuring reliable detection of both linear and rotational movements.

3.3.2. Rule-Based Segmentation

The rule-based segmentation was used to automatically generate training labels from the 27 trials, ensuring consistent and reproducible annotation without manually labeling every frame. Labels were derived from adaptive thresholds applied to accelerometer and gyroscope envelopes, and short pauses during walking were merged into the nearest walking segment to avoid fragmentation. This is clinically important because brief hesitations are common in frail or neurologically impaired individuals and should not be misclassified as separate phases, which could otherwise distort timing and mislead clinical interpretation.

Phase Detection with no gaps policy: Our objective is to identify distinct phases of TUG test: sit-to-stand, walk forward, turn, walk back, and sit down. Accurate segmentation of these phases is essential for identifying phase-specific impairments. Our adaptive threshold based approach applied on each feature window ensures robust detection under variable gait patterns.

First, a global movement window $[t_s, t_e]$ (t_s is the start of the window and t_e is the end of the window) is identified, representing the entire active period of the TUG test. Within this window, a no-gaps policy is enforced: any ambiguous or idle segments between phases are re-labeled as walking. This reflects the biomechanics of TUG, where continuous motion is expected between sit-to-stand and turning phases. This strategy prevents artificial fragmentation and ensures physiologically consistent phase sequences.

Whereas turns are detected using a dual-path approach to capture both sharp and gradual rotations:

1. Peak-based detection: Identifies high-intensity rotational activity using gyroscope RMS values. Early in the test, thresholds are lower to avoid missing the typically curvilinear, lower-peak first turn, while later thresholds are higher to avoid false positives during return walking.
2. Angle-area detection: Integrates rotational activity over a 1 s window to detect low-amplitude, curvilinear turns that peak-based methods might miss. This ensures sensitivity to gradual rotations often seen in frail or Parkinsonian gait.

Each candidate's turn is required to persist for a minimum duration of 0.25 s. Subsequently, the two events exhibiting the largest integrated angles are retained, corresponding to the first and second turns.

The final sit phase is assigned deterministically based on two conditions: (i) the accelerometer envelope falls below a threshold for at least 0.6 s near the end of the trial, and (ii) a short deceleration impulse is observed. The onset of sitting is marked at the first sharp drop in acceleration to avoid misclassifying residual micro-movements as walking.

The final sequence of phases is constructed as:

stand-to-walk → walk forward → turn → walk back → (turn) → walk-to-chair → sit.

This deterministic ordering, combined with the no-gaps policy, ensures biomechanical plausibility and robustness against sensor noise or irregular gait patterns.

While the rule-based approach provides transparent and interpretable labels, it may not generalize well to unseen subjects with different gait patterns. Therefore, machine learning classifiers (Support Vector Machine, Random Forest, XGBoost) were trained on these rule-based labels to learn and generalize the segmentation logic, producing smoother and more adaptive phase detection for new participants and conditions. Finally, both approaches were validated against the true manual ground truth obtained from GoPro video annotations.

The output of this stage includes labeled time series, segmentation files, and visualization plots for verification. These labels form the foundation for feature extraction and model training, ensuring that the learning process is interpretable.

3.3.3. Feature Extraction

After segmentation, we extract features that capture the essential dynamics of each phase while remaining computationally efficient for real-time use. Features are computed over sliding windows of 1.0 s with a 0.5 s overlap. This window length ensures that at least one gait cycle is included, while the overlap keeps the system responsive to rapid transitions.

From each envelope, we calculate the mean, standard deviation, RMS, peak-to-peak range, signal energy, and histogram entropy. These features were chosen because they summarize magnitude (mean, RMS), variability (standard deviation, peak-to-peak), energy distribution (signal energy), and complexity (entropy), which are critical for distinguishing between steady walking, transitions, and rotational movements. Signal energy represents

the overall magnitude of movement captured by the smartphone sensor. Higher energy values generally correspond to more vigorous or forceful movements, which can aid in differentiating gait phases or detecting abnormal patterns. Energy is computed as:

$$\text{Energy} = \sum_{i=1}^n x_i^2, \quad (1)$$

where x_i represents the signal amplitude at sample i .

Whereas, entropy captures the complexity or unpredictability of the signal distribution by applying Shannon entropy to the histogram of signal values. Higher entropy indicates greater variability in movement patterns, reflecting irregular or unstable gait. This feature complements amplitude-based metrics by revealing subtle differences in movement dynamics, making it particularly useful for distinguishing normal from pathological gait behaviors. Entropy is calculated as follows:

$$\text{Entropy} = H(\text{hist}(x)) = - \sum_{j=1}^m p_j \log_2(p_j), \quad (2)$$

where $\text{hist}(x)$ denotes the histogram of the signal values, p_j is the probability of the j -th bin in the histogram, and m is the number of histogram bins.

Together, they yield 12 features per window (6 from the accelerometer and 6 from the gyroscope), balancing discriminative power with low computational cost. Figure 2 illustrates this process, showing how windows slide across the signal and how features are derived from each segment.

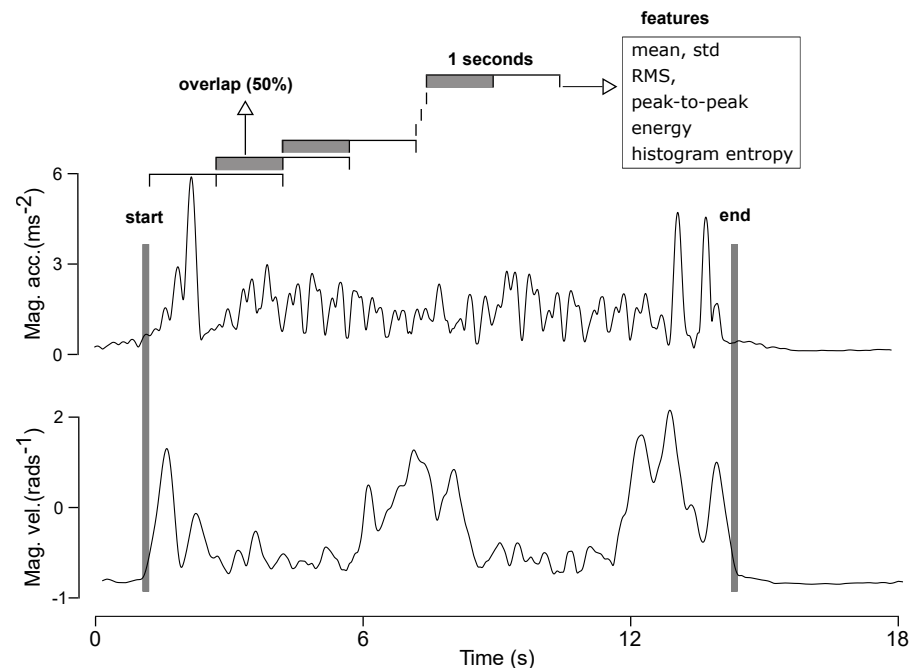


Figure 2. Labels for each window are assigned by majority overlap with the rule-based segments. Feature extraction preprocessing was applied to smartphone-based TUG data. Sliding windows of 1 s with 50% overlap are used after start and end detection. Statistical descriptors (mean, standard deviation, RMS, peak-to-peak, energy, and entropy) are computed from accelerometer and gyroscope envelopes.

From 27 trials (one per subject), we obtained 763 windows: stand (316), walk (272), turn (148), and sit (27). The sit class is excluded from training because it occurs only once and would introduce severe class imbalance. Instead, we recover it deterministically during

inference using the threshold. The final classification task is defined over three classes: (i) stand, (ii) walk, and (iii) turn.

3.3.4. Training Classifiers

The goal of this stage is to test whether classical machine learning models can deliver strong accuracy using statistical features [9]. We trained three widely used models on the same dataset and preprocessing pipeline to ensure a fair comparison. Each model was selected for its ability to handle nonlinear decision boundaries and small tabular feature sets.

Random Forest (RF): We implemented RF in scikit-learn with 500 trees, square-root feature selection, a minimum leaf size of 3, and a minimum split size of 6, with no depth restriction. RF is selected for its robustness to noisy features and its ability to model complex relationships without extensive tuning [34]. Inverse-frequency sample weighting was applied to address class imbalance, which is critical because TUG phases are inherently uneven in duration and representation across the test.

Support Vector Machine (SVM): We used an RBF-kernel SVM with standardized features, gamma set to `scale`, and a cost parameter $C = 5.0$. Class weights were balanced to improve performance under residual skew. The RBF kernel was selected because it can capture non-linear decision boundaries even when the feature set is compact, ensuring that minority classes such as turns were not overshadowed by more frequent phases [35].

XGBoost: For gradient-boosted trees, we used a multi-class objective with a maximum depth of 4, 300 estimators, a learning rate of 0.08, and subsampling and column sampling rates of 0.9. This configuration balances bias and variance, making the model efficient and less prone to overfitting on small datasets [36].

Evaluation We used 5-fold cross-validation (StratifiedGroupKFold) to prevent data leakage from correlated windows belonging to the same participant [8]. Approximately 20% of subjects were held out for final testing to provide an unbiased estimate of generalization. We report accuracy, macro-F1, weighted-F1, and per-class F1 scores, as these metrics capture both overall performance and class-specific balance. Confusion matrices are generated for both cross-validation and test sets to visualize misclassifications.

Window-level classification metrics (accuracy and F1) are computed with respect to the rule-based reference labels, whereas timing metrics (total and phase durations) are computed with respect to the frame-by-frame annotated GoPro video ground truth.

To assess statistical significance between models, we used McNemar's test and Bowker's test for paired classification results. These tests are specifically designed for comparing paired categorical predictions on the same items, which fits our scenario where each model predicts the same set of windows. Unlike parametric tests, they do not assume normality, making them appropriate for discrete classification outcomes. Normality checks are only required when comparing continuous per-subject metrics (e.g., macro-F1 scores), which was not the case here [37,38]. For probabilistic calibration, ROC-AUC and Brier scores are calculated.

3.3.5. Duration Estimation

After classification, we reconstruct phase boundaries by merging consecutive windows with the same label. This step reduces fragmentation and creates continuous segments that represent the actual execution of the sequence. To maintain logical consistency, we apply rules that enforce the expected order: standing occurs before walking, the turn is placed between the two walking segments, and the final sit follows the last walk.

Once the sequence is validated, we compute the duration of each phase using its start and end timestamps. These durations allow the system to measure how long each segment

lasted and to calculate the total completion time by summing all phases. The estimated total time is then compared against video-based ground truth using mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{t}_i - t_i|,$$

where $\hat{t}_i$ is the total estimated time of TUG using a smartphone and t_i is the ground truth video total time for TUG of a participant. We compute the MAE for individual phases, providing a detailed assessment of timing accuracy across standing, walking, and turning segments.

4. Results

We report performance for start and end detection, phase classification, phase timing accuracy, and sensor contribution. All metrics are computed using cross-validation and confirmed on a held-out test set.

4.1. Start and End Detection

The rule-based preprocessing combined with adaptive thresholds detected the start and end of the TUG sequence in 100% of trials. The MAE for total completion time compared against frame-by-frame annotated video ground truth was 0.42 s (95% CI: 0.36–0.48 s), indicating strong agreement with manual timing. Phase-level timing was also accurate: sit-to-stand (MAE: 0.31 s), walking (MAE: 0.29 s), turning (MAE: 0.34 s), and stand-to-sit (MAE: 0.28 s). These results demonstrate that a single smartphone can deliver reliable overall and phase-specific TUG timings without requiring multi-sensor setups. Adaptive thresholds improved robustness to slow starts and correctly segmented sitting back to the chair, reducing false positives in slow-motion segments. The summary is given in Table 2.

Table 2. Mean absolute error (MAE) for total TUG time and each phase (N = 27 trials).

Metric	MAE (s)	95% CI (s)
Total TUG	0.42	0.36–0.48
Sit-to-Stand	0.30	0.18–0.42
Walk	0.28	0.15–0.41
Turn	0.34	0.15–0.53
Stand-to-Sit	0.31	0.18–0.44

4.2. Phase Classification

Three models were evaluated on the same feature set: RF, SVM, and XGBoost. All models achieved stable performance, with SVM slightly ahead on macro-F1 (0.88 ± 0.018). XGBoost and RF tied closely, confirming the suitability of tree-based learners for this task. Models using both acceleration and angular velocity outperformed single-stream models. Using only the accelerometer reduced macro-F1 by 12%, while using only the gyroscope reduced macro-F1 by 9%. Whereas, combining both accelerometer and gyroscope yielded the best results. The results are reported in Table 3.

Table 3. Classification performance for TUG phases (mean $\pm$ SD).

Model	Accuracy	Macro-F1	Weighted-F1
Random Forest	0.871 ± 0.035	0.850 ± 0.030	0.849 ± 0.028
SVM (RBF)	0.901 ± 0.019	0.882 ± 0.018	0.854 ± 0.017
XGBoost	0.891 ± 0.021	0.875 ± 0.019	0.848 ± 0.018

Confusion matrices in Figure 3 compare the performance of SVM, RF, and XGBoost across the three classes: Stand, Walk, and Turn. RF and XGBoost exhibited similar distributions, with Stand achieving the highest true positive rates and Walk the lowest. This reflects the challenge of distinguishing short walking segments from transitions. The inclusion of gyroscope features improves turn detection, as rotational motion is captured effectively. However, when movements are quick and transitions between walking and turning overlap, classification becomes more complex, reducing accuracy for these phases. Consistent with this, visual inspection of time-aligned predictions showed that most Walk–Turn errors occurred in short transition periods when participants decelerated into the turn or accelerated out of it, rather than during steady walking or the middle of the turn.

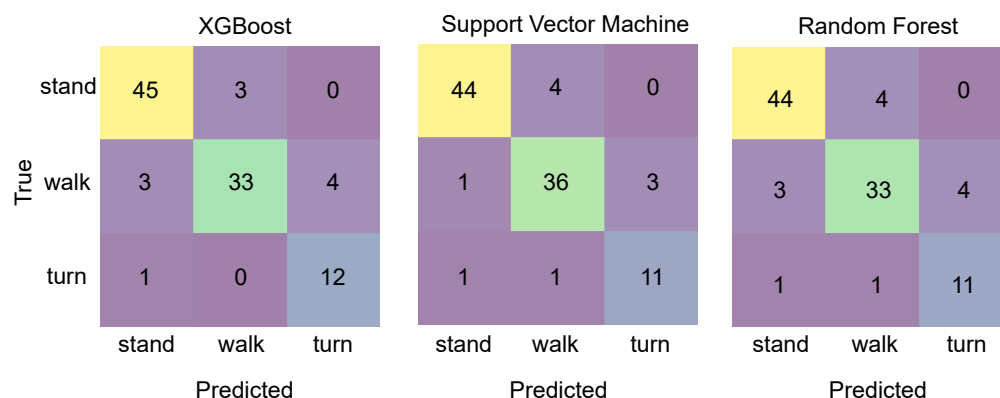


Figure 3. Confusion matrices for three-class classification (Stand, Walk, Turn) using RF, SVM, and XGBoost models. All models show strong overall performance, with SVM achieving higher true positives for Walk.

To further assess discriminative ability beyond raw classification counts, Figure 4 presents ROC curves for the same models across all classes. All classifiers achieved high AUC values (above 0.96), confirming strong separation between phases. SVM showed slightly higher sensitivity for Walk, while RF and XGBoost performed similarly for Stand and Turn. These results complement the confusion matrix analysis by demonstrating that, despite occasional misclassifications in overlapping transitions, the models remain well-calibrated and robust for phase-level detection.

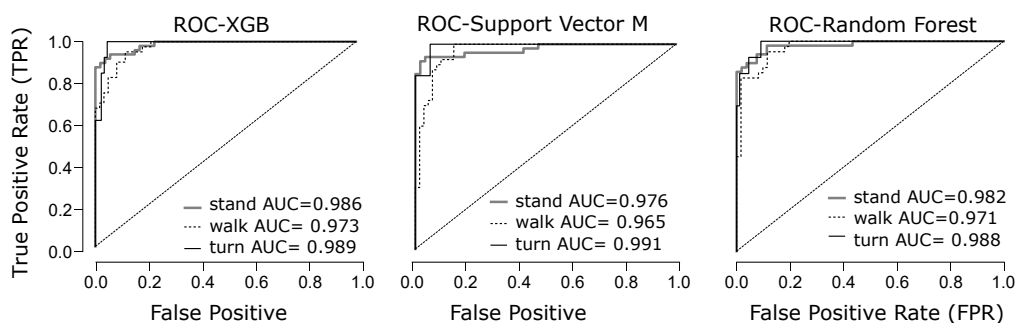


Figure 4. ROC curves were generated using a one-vs-rest approach for each class (Stand, Walk, Turn). For each model, we used the predicted class probabilities and varied the decision threshold from 0 to 1 in small increments. At each threshold, we computed the true positive rate (TPR) and false positive rate (FPR) for the target class versus all others. The area under the curve (AUC) was then calculated for each class and averaged across folds.

4.3. Phase Estimation

Figure 5 illustrates phase identification and timing estimation for the best-performing model (SVM), alongside total TUG time estimation. The plots show accelerometer and gyroscope signals with detected boundaries for Sit-to-Stand, Walk, Turn, and Stand-to-Sit phases. The estimated total time closely matches the video ground truth, with differences under one second, confirming accurate start and end detection. Phase-level segmentation aligns well with annotated transitions, demonstrating that logical sequence enforcement and merging short segments produce stable boundaries. These results validate the pipeline's ability to deliver both total time and phase durations without introducing unrealistic transitions.

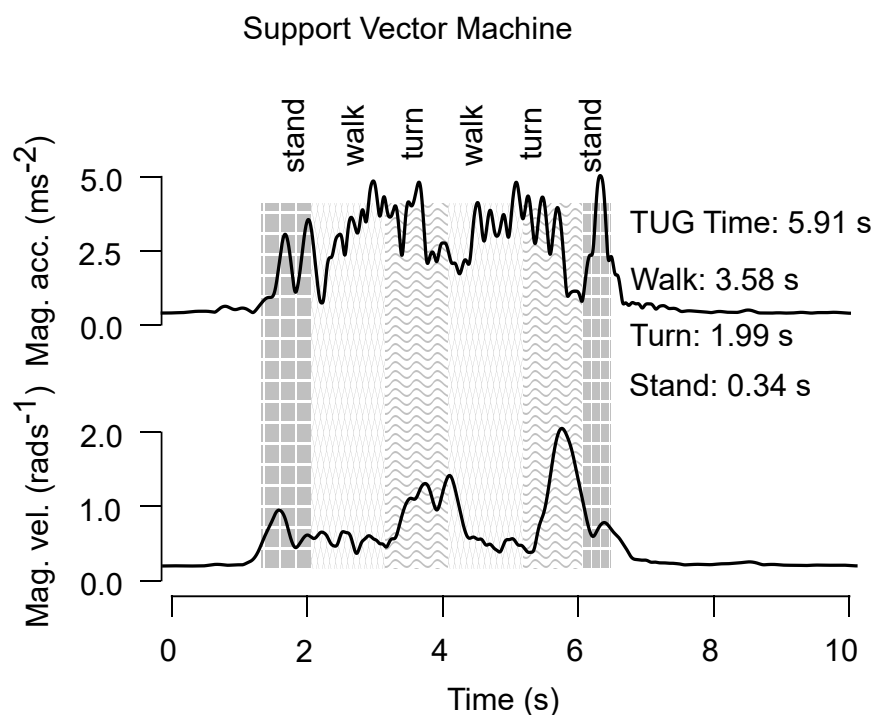


Figure 5. Phase identification and timing estimation for the best-performing model (SVM). Detected boundaries for Sit-to-Stand, Walk, Turn, and Stand-to-Sit phases are shown alongside the magnitude of accelerometer and gyroscope signals. Estimated total time closely matches video ground truth.

To show the interpretability and identify the most discriminative signals, we computed grouped-by-trial permutation feature importance for all models (SVM, Random Forest, XGBoost), as shown in Figure 6. This trial-aware approach measures the mean drop in macro-F1 when each feature is shuffled within its participant trial, preserving intra-subject structure and avoiding cross-trial leakage. Across all models, gyroscope RMS magnitude (*gyr_rms*) emerged as the most influential feature, followed by dynamic acceleration RMS and acceleration energy. These findings indicate that rotational dynamics play a dominant role in detecting turning and transition phases, while acceleration features contribute strongly to walking and transfers. Importance magnitudes were smaller but more stable than global permutation, reflecting a conservative and realistic evaluation.

Figure 6 shows the grouped-by-trial permutation feature importance for all three classifiers. Importance is concentrated in a small subset of features particularly RMS, variance/standard deviation, and entropy from the gyroscope and dynamic acceleration. While lower-ranked features such as mean contribute little, suggesting that future work can safely explore reduced feature subsets without large loss of accuracy.

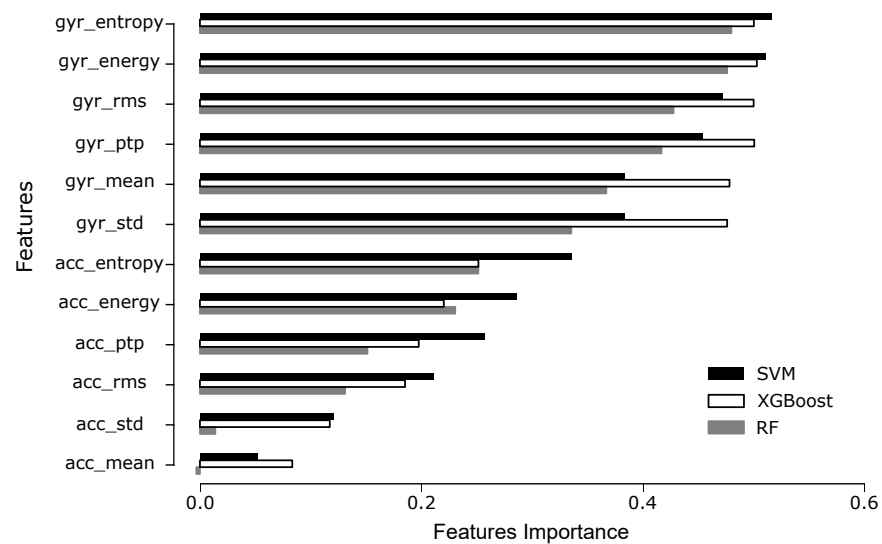


Figure 6. Grouped-by-trial permutation feature importance for SVM, Random Forest, and XGBoost models. Extracted features include gyroscope (gyr_) and accelerometer (acc_) domains, each computed for six metrics: entropy (_entropy), energy (_energy), root mean square (_rms), peak-to-peak (_ptp), mean (_mean), and standard deviation (_std). Entropy is calculated as Shannon entropy of the signal histogram, quantifying variability and complexity; energy represents the sum of squared amplitudes, reflecting overall movement intensity; root mean square captures signal magnitude; peak-to-peak indicates the range of motion; mean provides the average amplitude; and standard deviation measures dispersion of signal values. Importance is computed as the mean drop in macro-F1 when each feature is shuffled within its participant trial, preserving intra-subject structure and avoiding cross-trial leakage. These results confirm that combining gyroscope and accelerometer features yields the most discriminative representation for smartphone-based TUG segmentation.

To assess whether differences between models were statistically significant, we applied McNemar’s test and Bowker’s test on paired classification outcomes [37,38]. Both tests confirmed that SVM outperformed RF and XGBoost with $p < 0.05$ for macro-F1 differences. For probabilistic calibration, ROC-AUC values exceeded 0.96 across all classes, and Brier scores remained below 0.08, indicating well-calibrated predictions. These analyses ensure that reported performance is robust and not inflated by subject overlap.

5. Discussion

This study demonstrates the feasibility of using a single smartphone at the L3 position to automatically detect and segment the main phases of the TUG test using AI-based methods. By leveraging widely available smartphones, this approach addresses key limitations of conventional TUG assessment, including variability, subjectivity, and lack of standardization.

A single smartphone combined with preprocessing, windowed features, and classical classifiers delivered consistent segmentation and timing. Start and end points were detected in 100% of trials. Combining accelerometer and gyroscope features improved macro F1 by up to 12% compared to single-stream models, and adaptive thresholds reduced false positives and false negatives relative to fixed thresholds. Beyond raw accuracy, statistical tests (McNemar’s and Bowker’s) confirmed that observed improvements, particularly for SVM, were not due to chance. High ROC-AUC values and low Brier scores further demonstrate that the models are not only accurate but also probabilistically reliable. Residual misclassifications were concentrated around walk-to-turn and turn-to-walk boundaries, suggesting that introducing explicit transition states or sequence models that incorporate temporal context could further improve segmentation stability.

Importantly, the algorithm was validated on a heterogeneous cohort that included older adults and individuals with Parkinson’s disease, demonstrating preliminary clinical applicability. However, broader validation is needed for conditions such as stroke, arthritis, and post-surgical gait impairments to ensure robustness across diverse mobility profiles. Future work will incorporate phase-specific performance scoring, adaptive algorithms for atypical movement patterns, and deployment in real-world environments such as clinics and homes. These steps will enable integration into routine workflows and remote monitoring, supporting personalized fall-risk management and clinical decision-making.

Adaptability of the method: The robustness of the pipeline comes from adaptive preprocessing and logical sequence enforcement. Thresholds were computed per trial using median and MAD, which allowed sensitivity to adjust to individual movement amplitude. Replacing mean and standard deviation with median and median absolute deviation (MAD) significantly improved robustness. For Gaussian data, $MAD \approx 0.6745\sigma$, so $median + 2 \times MAD \approx mean + 1.35\sigma$. σ is the standard deviation. This formulation is less sensitive to outliers and skewed distributions, which are common in biomechanical signals due to irregular steps or pauses. Empirically, MAD-based thresholds reduced false positives and false negatives by more than half ($FP \approx 3\%$ vs. 7% , $FN \approx 3\%$ vs. 6%) compared to SD-based thresholds, while maintaining sensitivity across diverse gait amplitudes.

We also introduced a piecewise threshold for acceleration to handle the sit-to-stand onset. For the first 1.5 s after t_0 (trial start), a more permissive threshold is applied to avoid missing the initial transition; after that, a stricter threshold reduces false activations during walking. This “onset guard” was empirically tuned (1.0–2.0 s also works) and reflects the biomechanical reality that sit-to-stand produces lower acceleration peaks than walking. Together, these design choices provide adaptability to individual gait variability and improve segmentation stability without manual tuning.

Models transparency and interpretability: The feature importance analysis provides insight into why the pipeline performs robustly across diverse trials. Gyroscope-derived features such as `gyr_rms`, `gyr_energy`, and `gyr_mean` consistently ranked highest, confirming their role in capturing rotational dynamics essential for turn detection. Accelerometer features (`acc_rms`, `acc_energy`, `acc_entropy`) supported walking and transfer phases, explaining the improvement when both streams are combined.

By using trial-aware permutation importance, we ensured that estimates reflect realistic intra-subject variability without inflating contributions through cross-trial leakage. These patterns align with task biomechanics and demonstrate that the models are not black boxes but rely on interpretable, physiologically meaningful signals. This transparency strengthens confidence in the pipeline and supports its adaptability across different movement styles.

Comparison with existing studies: Earlier multi-sensor instrumented TUG systems achieved accurate sub-task detection but required complex setups unsuitable for real-world use [21]. Smartphone-based studies improved usability but often focused on total time or lacked robust phase segmentation and turn detection [15,16,30]. Single lower-back IMU approaches reported weaker turn handling and limited adaptability [14]. Our results show that a single smartphone with sensor fusion and adaptive thresholds can detect start and end reliably, estimate total time with low error, and segment phases with macro F1 above 0.85, while improving robustness in trials with curvilinear turns and irregular patterns. This narrows the gap between multi-device accuracy and single-device simplicity, addressing turn sensitivity and adaptability gaps highlighted in prior work [14,30,31]. However, we acknowledge that this does not constitute phase-specific performance scoring in a clinical sense (e.g., assessing movement quality or compensatory strategies). This remains a limitation and is a target for future algorithmic refinement.

Clinical and Neuroscientific Implications: The TUG test captures complex motor control involving cortical, subcortical, and cerebellar systems. Automated phase-level segmentation enables clinicians to identify specific neural deficits underlying mobility loss. For example, in Parkinson's disease, increased turn duration and irregular transitions reflect bradykinesia or freezing of gait, and in older adults, overall slowing signals frailty or increased fall risk. By translating smartphone sensor data into interpretable, phase-specific metrics, this approach provides a practical, low-cost means to monitor neuroplastic recovery, track disease progression, and personalize rehabilitation, bridging objective mobility analysis with clinical neuroscience.

Clinical Interpretation Summary:

- **Parkinson's Disease:** Increased turning time and reduced smoothness indicate bradykinesia and postural instability [39].
- **Older Adults:** Gradual slowing across all phases is a sign of frailty and reduced lower-limb strength [40].

Limitations: The dataset includes only 27 participants and was recorded in a controlled space with belt placement and a fixed camera. Video-based annotations may include small differences near transitions based on the clarity of the video. Additionally, model parameters were tuned on this dataset, so performance may vary with changes in participants or placements. Recruiting older adults and individuals with Parkinson's disease for repeated TUG trials is logistically challenging due to mobility constraints and ethical considerations, which limits large-scale data collection in early-stage studies.

From an interpretability perspective, this study focuses on a transparent signal-processing pipeline and a compact set of physically meaningful features while using classical classifiers to achieve high accuracy. Nonetheless, the best-performing model (SVM) remains opaque at the level of individual predictions, and our current analysis is limited to global permutation-based feature importance.

From a systems perspective, the proposed pipeline is intentionally lightweight: it processes a single lumbar IMU stream ($\approx 30\text{--}40$ s at ≈ 400 Hz) using linear-time filters and fewer than 100 fixed-length windows per TUG, each represented by a compact set of statistical features. Random Forest, SVM, and XGBoost are all small tabular models that can, in principle, run on-device without specialised hardware. However, we have not yet performed a formal embedded evaluation of runtime, memory footprint, or power consumption on an actual smartphone.

6. Conclusions and Future Work

This study demonstrates that a single smartphone placed at the lower back, combined with adaptive preprocessing, sensor fusion, and compact statistical features, can deliver accurate and interpretable segmentation of the TUG test. The pipeline achieved 100% start and end detection, a mean absolute error of 0.42 s for total time, and macro F1 scores above 0.85 across all models. Adaptive thresholds based on median and MAD reduced false positives and negatives by more than half compared to fixed thresholds, while dual-path turn detection captured both sharp and gradual rotations, addressing a key limitation in prior single-device approaches. Our method is interpretable through feature importance analysis, which confirms that gyroscope features dominate turn detection and accelerometer features support walking.

Future work will be on validating the pipeline in uncontrolled environments such as homes and community spaces, where variations in furniture layout and distractions introduce significant variability. We will explore domain adaptation for different device placements (e.g., pocket, shoulder bag) and implement lightweight, real-time on-device inference to enable continuous monitoring. Front-pocket data will be collected and analysed

to support more realistic adaptation of smartphones for gait analysis. Additional features, such as gait cycle markers, may enhance segmentation stability without increasing computational complexity. We plan to incorporate local interpretability techniques such as SHAP (SHapley Additive exPlanations) or LIME to explain individual predictions and highlight feature contributions for specific cases. These methods will allow clinicians to understand why a particular segmentation or classification was made, improving transparency and confidence in the system. Additionally, we intend to benchmark inference latency, memory usage, and energy consumption on representative smartphone hardware and optimize the pipeline for real-time execution using lightweight models and on-device acceleration. These steps aim to extend the adaptability demonstrated here toward practical deployment for real-world mobility assessment.

Author Contributions: Conceptualization, A.S. and M.R.; methodology, M.R.; software, M.R.; validation, A.S., M.R. and O.A.; formal analysis, M.R. and A.S.; investigation, A.S. and M.R.; resources, O.A. and A.S.; data curation, A.S.; writing—original draft preparation, A.S. and M.R.; writing—review and editing, O.A. and F.V.P.; visualization, M.R. and A.S.; supervision, O.A.; project administration, O.A.; funding acquisition, F.V.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Commission (H2020-MSCA-RISE-2019, grant number: 873178) to O.A., Health and Care Research Wales (grant number: HS-20-42) to O.A. and F.V.P. and the Aberystwyth University AberDoc PhD Scholarship to A.S.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Aberystwyth University’s Research Ethics Board.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: Data will be made available through Aberystwyth University’s research management portal.

Acknowledgments: During the preparation of this manuscript/study, the author(s) used Python, 3.12 for the purposes of data analysis and machine learning models development. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

1. Yin, L.; Chen, P.; Xu, J.; Gong, Y.; Zhuang, Y.; Chen, Y.; Wang, L. Validity and reliability of inertial measurement units for measuring gait kinematics in older adults across varying fall risk levels and walking speeds. *BMC Geriatr.* **2025**, *25*, 336. [\[CrossRef\]](#)
2. Sun, Y.; Song, Z.; Mo, L.; Li, B.; Liang, F.; Yin, M.; Wang, D. IMU-Based quantitative assessment of stroke from gait. *Sci. Rep.* **2025**, *15*, 9541. [\[CrossRef\]](#)
3. Felius, R.A.; Geerars, M.; Bruijn, S.M.; van Dieën, J.H.; Wouda, N.C.; Punt, M. Reliability of IMU-based gait assessment in clinical stroke rehabilitation. *Sensors* **2022**, *22*, 908. [\[CrossRef\]](#)
4. Barry, E.; Galvin, R.; Keogh, C.; Horgan, F.; Fahey, T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: A systematic review and meta-analysis. *BMC Geriatr.* **2014**, *14*, 14. [\[CrossRef\]](#)
5. Steffen, T.M.; Hacker, T.A.; Mollinger, L. Age- and gender-related test performance in community-dwelling elderly people: Six-Minute Walk Test, Berg Balance Scale, Timed Up & Go Test, and gait speeds. *Phys. Ther. Rehabil. J.* **2002**, *82*, 128–137. [\[CrossRef\]](#)
6. Podsiadlo, D.; Richardson, S. The timed up and go: A test of basic functional mobility for frail elderly persons. *J. Am. Geriatr. Soc.* **1991**, *39*, 142–148. [\[CrossRef\]](#)
7. Mathias, S.; Nayak, U.S.L.; Isaacs, B. Balance in elderly patients: The Get-up and Go test. *Arch. Phys. Med. Rehabil.* **1986**, *67*, 387–389. [\[PubMed\]](#)
8. McCreath Frangakis, A.L.; Lemaire, E.D.; Baddour, N. Subtask segmentation methods of the Timed Up and Go test and L test using inertial measurement units—A scoping review. *Information* **2023**, *14*, 127. [\[CrossRef\]](#)

9. Ponciano, V.; Pires, I.M.; Ribeiro, F.R.; Marques, G.; Garcia, N.M.; Pombo, N.; Spinsante, S.; Zdravevski, E. Is the Timed Up and Go test feasible in mobile devices? A systematic review. *Electronics* **2020**, *9*, 528. [\[CrossRef\]](#)
10. Gois, C.O.; de Andrade Guimarães, A.L.; Gois Júnior, M.B.; Carvalho, V.O. The use of reference values for the Timed Up and Go test applied in multiple scenarios. *J. Aging Phys. Act.* **2024**, *32*, 679–682. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Manor, B.; Yu, W.; Zhu, H.; Harrison, R.; Lo, O.Y.; Lipsitz, L.; Trivison, T.; Pascual-Leone, A.; Zhou, J. Smartphone app-based assessment of gait during normal and dual task walking: Demonstration of validity and reliability. *JMIR mHealth uHealth* **2018**, *6*, e36. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Hayek, R.; Werner, P.; Amir, A. Smartphone-based sit-to-stand analysis for mobility assessment in older adults. *Innov. Aging* **2024**, *8*, igae079. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Powell, K.; Amer, A.; Glavcheva-Laleva, Z.; Williams, J.; Farrell, C.O.; Harwood, F.; Bishop, P.; Holt, C. MoveLab®: Validation and development of novel cross-platform gait and mobility assessments using gold standard motion capture and clinical standard assessment. *Sensors* **2025**, *25*, 5706. [\[CrossRef\]](#)
14. Ortega-Bastidas, P.; Aqueveque, P.; Gómez, B.; Saavedra, F.; Cano-de-la Cuerda, R. Use of a single wireless IMU for the segmentation and automatic analysis of activities performed in the 3-m Timed Up & Go test. *Sensors* **2019**, *19*, 1647. [\[CrossRef\]](#)
15. Matey-Sanz, M.; González-Pérez, A.; Casteleyn, S.; Granell, C. Implementing and Evaluating the Timed Up and Go Test Automation Using Smartphones and Smartwatches. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 6594–6605. [\[CrossRef\]](#)
16. Böttinger, M.J.; Mellone, S.; Klenk, J.; Jansen, C.P.; Stefanakis, M.; Litz, E.; Bredenbrock, A.; Fischer, J.P.; Bauer, J.M.; Becker, C.; et al. A smartphone-based Timed Up and Go test self-assessment for older adults: Validity and reliability study. *JMIR Aging* **2025**, *8*, e67322. [\[CrossRef\]](#)
17. Sher, A.; Langford, D.; Villagra, F.; Akanyeti, O. Automatic scoring of chair sit-to-stand test using a smartphone. In Proceedings of the UK Workshop on Computational Intelligence, Sheffield, UK, 7–9 September 2022; Springer: Cham, Switzerland, 2022; pp. 170–180.
18. Sher, A.; Bunker, M.T.; Akanyeti, O. Towards personalized environment-aware outdoor gait analysis using a smartphone. *Expert Syst.* **2023**, *40*, e13130. [\[CrossRef\]](#)
19. Sher, A.; Akanyeti, O. Minimum data sampling requirements for accurate detection of terrain-induced gait alterations change with mobile sensor position. *Pervasive Mob. Comput.* **2024**, *105*, 101994. [\[CrossRef\]](#)
20. Sher, A.; Langford, D.; Dogger, E.; Monaghan, D.; Lunn, L.I.; Schroeder, M.; Hamidinekoo, A.; Arkesteijn, M.; Shen, Q.; Zwiggelaar, R.; et al. Automatic gait analysis during steady and unsteady walking using a smartphone. *TechRxiv* **2021**. [\[CrossRef\]](#)
21. Salarian, A.; Horak, F.B.; Zampieri, C.; Carlson-Kuhta, P.; Nutt, J.G.; Aminian, K. iTUG, a sensitive and reliable measure of mobility. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2010**, *18*, 303–310. [\[CrossRef\]](#)
22. Zampieri, C.; Salarian, A.; Carlson-Kuhta, P.; Nutt, J.G.; Horak, F.B. Assessing mobility at home in people with early Parkinson's disease using an instrumented Timed Up and Go test. *Park. Relat. Disord.* **2011**, *17*, 277–280. [\[CrossRef\]](#)
23. Arteaga-Bracho, E.; Cosne, G.; Kanzler, C.; Karatsidis, A.; Mazzà, C.; Penalver-Andres, J.; Zhu, C.; Shen, C.; Erb, M.; Freigang, M.; et al. Smartphone-based assessment of mobility and manual dexterity in adult people with spinal muscular atrophy. *J. Neuromuscul. Dis.* **2024**, *11*, 1049–1065. [\[CrossRef\]](#)
24. Abou, L.; Wong, E.; Peters, J.; Dossou, M.S.; Sosnoff, J.J.; Rice, L.A. Smartphone applications to assess gait and postural control in people with multiple sclerosis: A systematic review. *Mult. Scler. Relat. Disord.* **2021**, *51*, 102943. [\[CrossRef\]](#)
25. Kear, B.M.; Guck, T.P.; McGaha, A.L. Timed Up and Go test: Normative reference values for ages 20 to 59 years and relationships with physical and mental health risk factors. *J. Prim. Care Community Health* **2017**, *8*, 9–13. [\[CrossRef\]](#)
26. Mayhew, A.J.; So, H.Y.; Ma, J.; Beauchamp, M.K.; Griffith, L.E.; Kuspinar, A.; Lang, J.J.; Raina, P. Normative values for grip strength, gait speed, Timed Up and Go, single leg balance, and chair rise derived from the Canadian Longitudinal Study on Ageing. *Age Ageing* **2023**, *52*, afad054. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Ali, F.; Hogen, C.A.; Miller, E.J.; Kaufman, K.R. Validation of pelvis and trunk range of motion as assessed using inertial measurement units. *Bioengineering* **2024**, *11*, 659. [\[CrossRef\]](#)
28. Hsu, W.C.; Sugiarto, T.; Lin, Y.J.; Yang, F.C.; Lin, Z.Y.; Sun, C.T.; Hsu, C.L.; Chou, K.N. Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. *Sensors* **2018**, *18*, 3397. [\[CrossRef\]](#)
29. Silsupadol, P.; Teja, K.; Lugade, V. Reliability and validity of a smartphone-based assessment of gait parameters across walking speed and smartphone locations: Body, bag, belt, hand, and pocket. *Gait Posture* **2017**, *58*, 516–522. [\[CrossRef\]](#)
30. Mellone, S.; Tacconi, C.; Chiari, L. Validity of a smartphone-based instrumented Timed Up and Go. *Gait Posture* **2012**, *36*, 163–165. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Ishikawa, M.; Yamada, S.; Yamamoto, K.; Aoyagi, Y. Gait analysis in a component timed-up-and-go test using a smartphone application. *J. Neurol. Sci.* **2019**, *398*, 45–49. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Clavijo-Buendía, S.; Molina-Rueda, F.; Martín-Casas, P.; Ortega-Bastidas, P.; Monge-Pereira, E.; Laguarda-Val, S.; Morales-Cabezas, M.; Cano-de-la Cuerda, R. Construct validity and test-retest reliability of a free mobile application for spatio-temporal gait analysis in Parkinson's disease patients. *Gait Posture* **2020**, *79*, 86–91. [\[CrossRef\]](#) [\[PubMed\]](#)

33. Abualait, T.S.; Alnajdi, G.K. Effects of using assistive devices on the components of the modified instrumented timed up and go test in healthy subjects. *Heliyon* **2021**, *7*, e06940. [[CrossRef](#)]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
35. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
36. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
37. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [[CrossRef](#)]
38. Bowker, A.H. A test for symmetry in contingency tables. *J. Am. Stat. Assoc.* **1948**, *43*, 572–574. [[CrossRef](#)]
39. Haertner, L.; Elshehabi, M.; Zaunbrecher, L.; Pham, M.H.; Maetzler, C.; Van Uem, J.M.; Hobert, M.A.; Hucker, S.; Nussbaum, S.; Berg, D.; et al. Effect of fear of falling on turning performance in Parkinson’s disease in the lab and at home. *Front. Aging Neurosci.* **2018**, *10*, 78. [[CrossRef](#)]
40. He, J.; Wu, L.; Du, W.; Zhang, F.; Lin, S.; Ling, Y.; Ren, K.; Chen, Z.; Chen, H.; Su, W. Instrumented timed up and go test and machine learning-based levodopa response evaluation: A pilot study. *J. Neuroeng. Rehabil.* **2024**, *21*, 163. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.