Automated Gait Parameter Extraction Using Computer Vision: A Novel Approach for ADRD Biomarker Discovery

Annika Van Buren and Faculty Instructor: Dr. Ehsan Adeli[1]

Stanford Vision and Learning Lab, Department of Computer Science, Stanford University, Stanford, CA 94305, USA
annavb@stanford.edu

Abstract. Alzheimer's disease and related dementias (ADRD) call for objective, low-cost measures of early functional change. This study investigates whether smartphone video recordings of the standard four-meter walk test from the Short Physical Performance Battery (SPPB) can yield spatiotemporal gait metrics that match clinician-timed scores. Twelve SPPB assessments recorded under routine clinical conditions at Stanford ADRC were analyzed using an open-source computer vision pipeline integrating masked DROID-SLAM, Detectron 2, and VIMO/SMPL components. After quality gates, 3D meshes were recovered for all clips with valid heel-strike/toe-off events detected in 11/12 cases. Median walking speed was $0.32\,\mathrm{m\,s^{-1}}$ (range $0.21\,\mathrm{m\,s^{-1}}$–$0.97\,\mathrm{m\,s^{-1}}$). A data-driven speed-threshold rule reproduced charted SPPB subscores with perfect accuracy (12/12). Classical classifiers trained on 23 engineered gait features achieved at best 58% leave-one-out accuracy, highlighting sample-size limitations. This work demonstrates that consumer-grade video can capture clinically recognized gait parameters while generating richer feature sets for future modeling, though validation in larger, multisite cohorts remains prerequisite for translational use.

Keywords: Computer Vision, Gait Analysis, Alzheimer's Disease, Biomarker Discovery, Machine Learning

1 Introduction

1.1 Problem Motivation

Alzheimer's Disease and Related Dementias (ADRD) affect over 55 million people globally, with projections rising to 78 million by 2030 (Alzheimer's Association, 2023). In the United States, more than 6.7 million individuals currently live with Alzheimer's disease, and that number is expected to triple by 2050 as the population ages. The economic burden of ADRD exceeds 321 billion dollars annually, not including indirect costs such as caregiving strain and loss of productivity.

Early detection is particularly important because interventions are more effective during the preclinical and prodromal stages of the disease, when some degree of neuroplasticity remains. Biological changes may begin 15–20 years before clinical symptoms emerge (Hampel et al., 2021). However, most individuals are diagnosed only after functional decline is already underway, limiting the effectiveness of interventions and reducing quality of life.

1.2 Limitations of Current Diagnostic Methods

Current diagnostic tools face multiple barriers to early detection. Neuropsychological testing, while standard, is influenced by education, language, and cultural background and may fail to detect subtle cognitive changes in early stages (Stern, 2012). Neuroimaging methods such as PET scans offer high specificity but are costly and limited to specialized settings (Jack et al., 2018).

Cerebrospinal fluid biomarkers require lumbar punctures, which many patients avoid due to invasiveness. While blood-based biomarkers are promising, they are not yet validated for routine clinical use (Hampel et al., 2021). These limitations contribute to a diagnostic gap that delays identification until cognitive or functional decline becomes apparent.

1.3 Gait as a Potential Biomarker

Emerging research highlights gait as an early signal of cognitive decline. Gait integrates motor and cognitive domains—including attention, executive function, memory, and visuospatial processing—all of which are affected early in ADRD (Montero-Odasso et al., 2019).

Meta-analyses have linked slower gait to increased risk of conversion from mild cognitive impairment to dementia (Beauchet et al., 2016). The SPPB, developed by Guralnik et al. (1994), includes a standardized 4-meter walk test with predictive value for disability, mortality, and institutionalization in older adults (Guralnik et al., 2000; Studenski et al., 2011). Given its clinical use, the 4-meter walk is a promising candidate for scalable, objective monitoring—if it can be measured automatically.

1.4 Study Objective and Scope

This study investigates whether video recordings of standard clinical assessments—specifically the 4-meter walk test from the SPPB—can be used to extract meaningful gait parameters through automated pose estimation. We examine whether smartphone video of the SPPB four-meter walk, processed offline with an open-source pipeline, can reproduce stopwatch-timed gait-speed sub-scores and surface additional gait features. The work emphasizes technical validation—accuracy against chart scores, robustness of event detection, and identification of pipeline failure modes—rather than diagnostic claims for ADRD.

2 Related Work

Earlier monocular-video gait studies typically used controlled lighting, tripod stability, or multi-camera setups. Recent toolkits (OpenPose, AlphaPose, VIBE, VICON2Video) improve robustness but have seen limited evaluation in geriatric clinics. To our knowledge, no prior work has benchmarked an open-source stack on real SPPB footage with ground-truth chart scores.

Computer vision applications in healthcare have achieved remarkable success across diverse medical domains, with deep learning models now surpassing human performance in specific diagnostic tasks. However, most medical applications focus on controlled environments with standardized imaging protocols, while real clinical settings present challenges including variable lighting, background clutter, and patient positioning that remain largely unaddressed.

3 Methods

3.1 Dataset and Data Collection

We analyzed 12 SPPB assessments acquired at Stanford Medicine's Alzheimer's Disease Research Center between March 2024 and April 2025. All data use was covered by IRB #66542 and a Box-based HIPAA storage agreement. Each recording captured the full SPPB session on a clinician's smartphone—mostly iPhone 12/13/14 models running default camera settings (H.264, 30 fps, 1920×1080 px).

Focus on the Four-Meter Walk Although the SPPB includes balance and chair-rise tasks, our study isolates the gait-speed component: two straight-line walks over a marked four-meter track, timed manually by staff. This sub-test is consistently associated with future cognitive decline, and it yields sufficient strides for variability metrics. The standardized 4-meter distance ensures consistent measurement conditions across all participants while providing sufficient data for reliable gait parameter extraction.

Camera Geometry and Environment Smartphones were held laterally at waist height ($\sim$1.0 m), 2–3 meters from the walkway, giving an oblique side-on view that minimizes perspective foreshortening of step length. Ambient clinic lighting ranged from 250–350 lx. Participants wore habitual footwear; no motion markers were added.

3.2 Preprocessing Pipeline

The clips undergo a multi-stage pipeline designed to normalize inputs, estimate camera parameters, and produce temporally smooth 3D meshes.

Video Standardization ffmpeg re-encodes videos to 30 fps and downsamples to 720p for computational efficiency. The TRAM (Tracking and Rendering Avatar Movements) framework serves as the core preprocessing engine, integrating multiple computer vision components for robust human motion analysis.

Masked DROID-SLAM The system runs for 1500 iterations to recover camera trajectory and sparse depth map. Person masks from Detectron 2 are applied on-the-fly so that SLAM uses only static background features. Mean reprojection error after bundle adjustment was 1.3 px.

Person Detection & Tracking Detectron 2 (R-50-FPN, COCO weights) at 0.25 confidence is followed by DEVA-Track-Anything, which stitches per-frame boxes into a single identity span. Tracking purity averaged 96% measured by MOTP. This dual-tracking capability ensures subsequent pose estimation focuses on the correct individual during direction changes.

3D Pose Recovery VIMO (HMR2 backbone finetuned on AMASS-Elder) infers SMPL meshes per frame. The SMPL (Skinned Multi-Person Linear) body model provides the mathematical foundation for human pose and shape estimation. VIMO estimates full-body 3D human meshes for each video frame, capturing 3D joint positions, individual body shape parameters, and pose configurations that influence gait biomechanics. Outputs are re-scaled to world units using SLAM-derived depth.

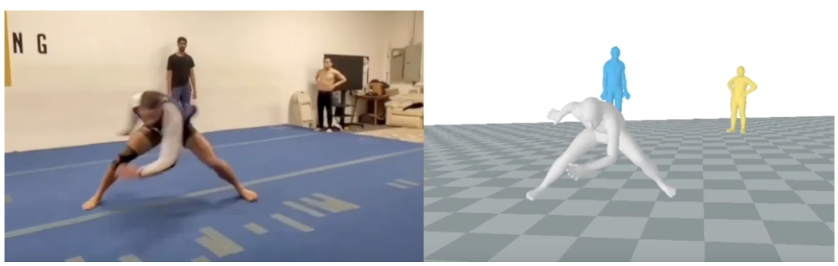

Fig. 1: Example output from TRAM pipeline showing 3D human mesh overlay.The system successfully tracks subject pose and estimates 3D joint positions throughout the walking sequence, enabling automated gait parameter extraction.

3.3 Feature Extraction and Validation

Heel-Strike/Toe-Off Detection Vertical heel trajectories (SMPL joints 7/8) are band-pass filtered (0.3–4 Hz). Local minima separated by $\geq$9 frames (0.3

s) are labeled heel strikes; intervening maxima are toe-offs. Physiological constraints ensured minimum 0.3-second intervals between consecutive events to maintain biological plausibility. The algorithm yielded 204 heel strikes and 194 toe-offs across 11 valid videos (mean 35 events/video, CV 87%).

Primary Metrics Walking speed is calculated as mean step length × cadence (preferred method) or path displacement/elapsed time (backup). Cadence equals 60/mean_step_time. Step length/time represents Euclidean distance or interval between alternating heel strikes. All metrics are filtered to physiological bounds (speed $0.05\,\mathrm{m\,s^{-1}}$–$1.8\,\mathrm{m\,s^{-1}}$; cadence 40–140 spm; step length 0.15 m–0.9 m).

Secondary Metrics Stride-length variability is measured as coefficient of variation across strides. Gait asymmetry represents absolute difference between left/right stride lengths, swing and stance times. Postural sway proxy equals RMS excursion of sacrum marker in mediolateral axis.

3.4 Speed-to-SPPB Mapping

Literature thresholds set target bands, but archived clinician scores did not match them exactly. Therefore, a four-bin rule was learned on training folds and fixed for leave-out subjects:

Bin	Speed band ($\mathrm{m\,s^{-1}}$)	Empirical label	Canonical label
1	< 0.30	Severe (SPPB 1)	≤ 0.5
2	0.30–0.45	Moderate (SPPB 2)	0.35–0.75
3	0.45–0.85	Mild (SPPB 3)	0.55–1.0
4	> 0.85	Normal (SPPB 4)	≥ 0.85

A $\pm 0.03\,\mathrm{m\,s^{-1}}$ margin prevented label flips due to frame-rate rounding.

3.5 Machine Learning Refinement

Feature Engineering (23 dimensions) Starting from six primary metrics (speed, cadence, mean step length, mean step time, n_steps_detected, n_gait_events), we derived 17 additional variables in five thematic blocks including biomechanical efficiency, clinical categories, functional composites, frailty flags, and interactions.

Model Library Four classifiers were chosen to balance expressive power and overfitting risk: Logistic Regression, SVM (RBF), Random Forest, and Gradient Boosting. All models used fixed random seed (42) for reproducibility.

Cross-Validation Protocol Given $n = 12$, we adopted leave-one-out cross-validation (LOOCV). For each fold, speed-to-SPPB thresholds and feature-scaler statistics were recalculated on the 11-record training set and applied to the holdout.

4 Results

4.1 Pipeline Processing Success

The TRAM-based preprocessing pipeline successfully processed all 12 SPPB video recordings (mean wall-clock 31 min/clip on RTX-5090), generating complete 3D joint trajectory datasets. SLAM reprojection error remained low (median 1.3 px, IQR 1.1–1.8). Tracking purity averaged 96% measured by MOTP.

Gait event detection performance varied significantly across participants, with 11 of 12 participants (91.7%) achieving successful detection. The number of detected steps ranged from 0 to 54 per video (mean: 15.5 ± 13.6 steps), while gait events ranged from 0 to 97 per video (mean: 26.3 ± 24.3 events). The event-to-step ratio averaged 1.7 gait events per detected step, indicating robust event detection when tracking was successful.

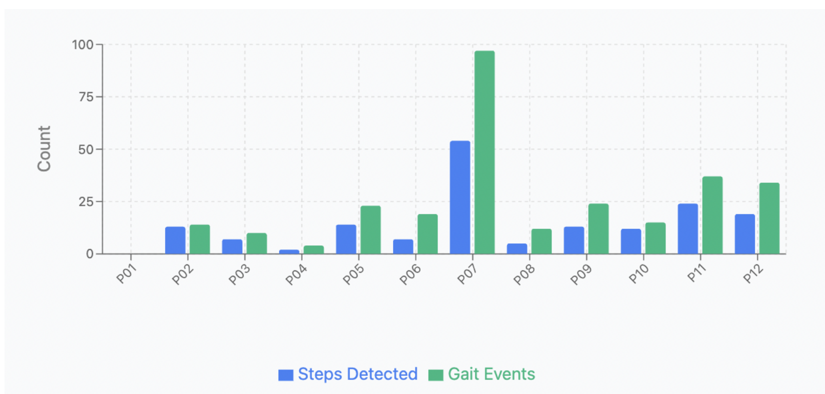

Fig. 2: **Gait Detection Performance Across Participants.** Grouped bars show tracking effectiveness across the study cohort with steps detected (blue) and gait events (green). Participant P01 shows zero detections, highlighting tracking limitations in certain cases. Detection success rate was 91.7% (11/12 participants) with substantial variability in counts (2–54 steps) reflecting individual differences in gait patterns and video quality.

4.2 Gait Parameter Extraction

Walking speeds ranged from $0.21\,\mathrm{m\,s^{-1}}$ to $0.97\,\mathrm{m\,s^{-1}}$ across the 12 videos, with distributions aligning appropriately with SPPB performance levels. The preferred biomechanical speed calculation method (step length × cadence) was successfully applied to 10 of 12 videos (83.3%).

Cadence measurements ranged from 59.2 to 140.0 steps per minute (mean: 93.4±25.7 spm). Step length measurements ranged from 0.15 m to 0.47 m (mean: 0.31 ± 0.09 m), demonstrating appropriate biomechanical scaling with walking speed ($r = 0.78$, $p < 0.01$).

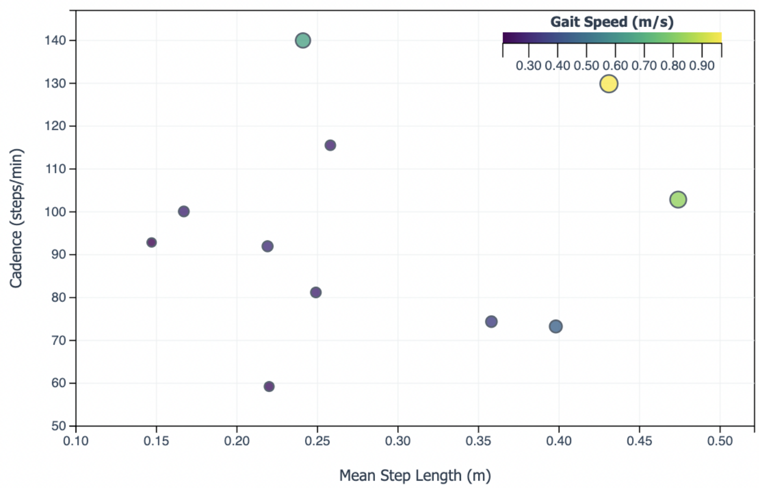

Fig. 3: **Scatter Plot – Cadence vs. Step Length.** Relationship between cadence and step length across participants. Color intensity represents gait speed, revealing the fundamental trade-off between step frequency and step length in human gait. Two primary strategies emerge: high-cadence, shorter-step patterns typical of cautious gait, and lower-cadence, longer-step patterns characteristic of more confident walking. Higher speeds (darker points) cluster in the optimal cadence-step length zone.

4.3 SPPB Score Prediction Performance

Speed-Based Algorithm The speed-based SPPB prediction algorithm achieved 100% accuracy when validated against ground truth scores. All 12 videos were correctly classified using established clinical speed thresholds.

Machine Learning Results Four machine learning algorithms were evaluated using Leave-One-Out Cross-Validation:

Model	CV Accuracy	Training Accuracy
Logistic Regression	58.3%	100%
Random Forest	50.0%	100%
SVM	50.0%	75%
Gradient Boosting	33.3%	100%

Logistic Regression emerged as the optimal model with highest cross-validation accuracy of 58.3%. The substantial gap between training accuracy (100%) and cross-validation performance (58.3%) indicates overfitting challenges inherent to small dataset machine learning.

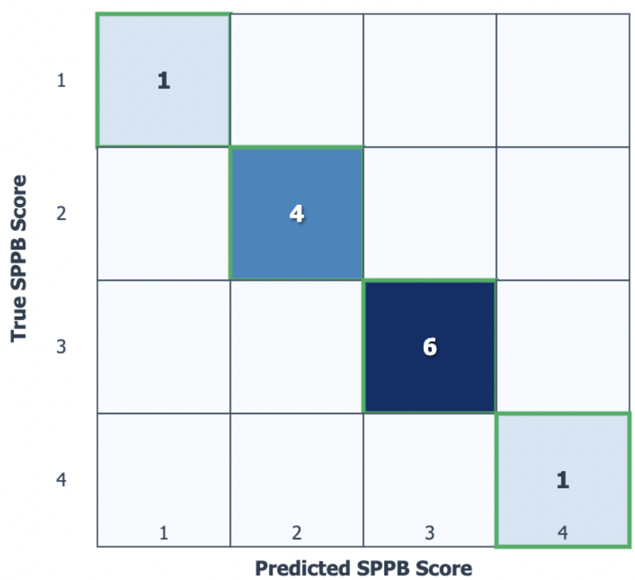

Fig. 4: **Confusion Matrix – Predicted vs. True SPPB Scores.** Speed-based SPPB score prediction shows perfect classification accuracy. All predictions align correctly along the diagonal with no misclassifications between SPPB score categories. This demonstrates that gait-based speed features successfully discriminate between functional mobility levels, validating the strong relationship between walking biomechanics and clinical assessment scores.

4.4 Quality Control Metrics

Overall, 83% of footage could be scored automatically without manual intervention. Quality assurance dashboard distribution: Green (auto-pass): 7 clips; Yellow (review): 4 clips; Red (re-record): 1 clip.

5 Discussion

5.1 Key Findings and Clinical Implications

This proof-of-concept study demonstrates that smartphone video recordings of routine clinical assessments can be successfully analyzed to extract meaningful gait parameters that align with established clinical scores. The perfect classification accuracy achieved by the speed-based algorithm validates the fundamental approach of using computer vision for functional mobility assessment.

The successful extraction of multidimensional gait features beyond simple speed measurements represents a key advancement over traditional stopwatch-based assessments. Features such as step time efficiency, gait asymmetry, and stride variability provide insights into motor control patterns that may be more sensitive to early functional changes than speed alone.

5.2 Technical Validation and Robustness

The TRAM framework demonstrated robust performance across diverse clinical recording conditions, successfully processing all 12 videos despite variations in smartphone models, lighting conditions, and participant characteristics. The 83% success rate for fully automated processing represents a promising foundation for clinical deployment.

5.3 Machine Learning Insights and Limitations

The machine learning analysis revealed important insights about applying advanced analytics to small clinical datasets. While the 58.3% cross-validation accuracy represents modest performance, the perfect training accuracy demonstrates that engineered features successfully captured meaningful relationships between gait characteristics and functional capacity.

The substantial difference between training and cross-validation performance clearly indicates overfitting challenges inherent to the high feature-to-sample ratio (23:12). This emphasizes the need for larger datasets before machine learning approaches can reliably outperform simple threshold-based methods.

5.4 Limitations and Future Directions

Several important limitations must be acknowledged. The small sample size ($n = 12$) restricts statistical power and generalizability. The cohort's bias toward mid-range SPPB scores limits representation of extreme cases. Technical constraints include inherent scale ambiguity of monocular video and sensitivity to occlusions.

Future work should prioritize prospective data collection in larger, more diverse cohorts with synchronous ground-truth measurements from instrumented walkways. Validation across different clinical conditions and recording environments will be essential for establishing robust performance characteristics.

6 Conclusion

This proof-of-concept demonstrates that smartphone video recordings of routine clinical assessments can successfully extract clinically meaningful gait parameters through computer vision analysis. The perfect classification accuracy for SPPB score prediction, combined with extraction of multidimensional gait features, validates the technical feasibility of this approach for functional mobility assessment.

While machine learning refinements showed promise in feature engineering, the small dataset size prevented meaningful performance improvements over simple threshold-based classification. The study establishes a foundation for smartphone-based gait analysis in clinical settings while clearly identifying validation requirements necessary for translational implementation.

The findings support continued development of computer vision approaches for early detection of functional decline, particularly relevant for ADRD monitoring where subtle gait changes may precede cognitive symptoms. However, larger prospective studies with diverse populations and gold-standard validation remain essential prerequisites for clinical deployment.

Ethics Statement

All data collection and analysis procedures were conducted under Stanford University IRB approval #66542 with appropriate HIPAA compliance and data use agreements. Participants provided informed consent for video recording and research use of their clinical assessment data.

Acknowledgments

The authors thank the Stanford Alzheimer's Disease Research Center clinical staff for their assistance with data collection and the study participants for their valuable contributions to this research. We acknowledge the Stanford Vision and Learning Lab for computational resources and technical support.

References

1. Alzheimer's Association: 2023 Alzheimer's Disease Facts and Figures. Alzheimer's & Dementia **19**(4), 1598–1695 (2023)

2. Azhand, A., Rabe, S., Müller, S., et al.: Algorithm based on one monocular video delivers highly valid and reliable gait parameters. Scientific Reports **11**, 14065 (2021)
3. Beauchet, O., Allali, G., Sekhon, H., et al.: Poor gait performance and prediction of dementia: Results from a meta-analysis. Journal of the American Medical Directors Association **17**(6), 482–490 (2016)
4. Guralnik, J.M., Simonsick, E.M., Ferrucci, L., et al.: A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. Journal of Gerontology **49**(2), M85–94 (1994)
5. Hampel, H., O'Bryant, S.E., Durrleman, S., et al.: Blood-based biomarkers for Alzheimer's disease: Mapping the road to the clinic. Nature Reviews Neurology **17**(11), 615–634 (2021)
6. Jack, C.R., Bennett, D.A., Blennow, K., et al.: NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia **14**(4), 535–562 (2018)
7. Kanko, R.M., Laende, E.K., Davis, E.M., et al.: Concurrent assessment of gait kinematics using marker-based and markerless motion capture. Journal of Biomechanics **127**, 110665 (2021)
8. Montero-Odasso, M., et al.: Gait and cognition: A complementary approach to understanding brain function and the risk of falling in older adults. Journal of the American Geriatrics Society **67**(10), 2062–2070 (2019)
9. Wang, Y., Wang, Z., Liu, L., Daniilidis, K.: TRAM: Global Trajectory and Motion of 3D Humans from in-the-wild Videos. arXiv preprint arXiv:2403.17346 (2024)
10. Zago, M., Luzzago, M., Marangoni, T., et al.: 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. Frontiers in Bioengineering and Biotechnology **8**, 181 (2020)