POLITECNICO DI TORINO
Repository ISTITUZIONALE

Continuous assessment of daily-living gait using self-supervised learning of wrist-worn accelerometer data

Original
Continuous assessment of daily-living gait using self-supervised learning of wrist-worn accelerometer data / Brand, Y.E., Buchman, A.S., Kluge, F., Palmerini, L., Becker, C., Cereatti, A., Maetzler, W., Vereijken, B., Yarnall, A.J., Rochester, L., Del Din, S., Mueller, A., Hausdorff, J.M., Perlman, O.r.. - In: NPJ DIGITAL MEDICINE. - ISSN 2398-6352. - 9:1(2026). [10.1038/s41746-026-02528-2]

Availability:
This version is available at: 11583/3015518 since: 2026-09-15T12:55:32Z

Publisher:
Nature

Published
DOI:10.1038/s41746-026-02528-2

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

07 November 2026

ARTICLE IN PRESS

npj Digital Medicine

https://doi.org/10.1038/s41746-026-02528-2

Article in Press

Continuous assessment of daily-living gait using self-supervised learning of wrist-worn accelerometer data

Received: 20 May 2025

Accepted: 27 February 2026

Cite this article as: Brand, Y.E., Buchman, A.S., Kluge, F. *et al.* Continuous assessment of daily-living gait using self-supervised learning of wrist-worn accelerometer data. *npj Digit. Med.* (2026). https://doi.org/10.1038/s41746-026-02528-2

Yonatan E. Brand, Aron S. Buchman, Felix Kluge, Luca Palmerini, Clemens Becker, Andrea Cereatti, Walter Maetzler, Beatrix Vereijken, Alison J. Yarnall, Lynn Rochester, Silvia Del Din, Arne Mueller, Jeffrey M. Hausdorff & Or Perlman

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

ARTICLE IN PRESS

Continuous Assessment of Daily-Living Gait Using Self-Supervised Learning of Wrist-Worn Accelerometer Data

Yonatan E. Brand (1,2), Aron S. Buchman (3), Felix Kluge (4), Luca Palmerini (5), Clemens Becker (6,7), Andrea Cereatti (8), Walter Maetzler (9), Beatrix Vereijken (10), Alison J. Yarnall (11,12,13), Lynn Rochester (11,12,13), Silvia Del Din (11,13), Arne Mueller (4), Jeffrey M. Hausdorff (2,14,15,16), Or Perlman* (1,15)

1. School of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel
2. Center for the Study of Movement, Cognition and Mobility, Neurological Institute, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel
3. Rush Department of Neurological Sciences, Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
4. Biomedical Research, Novartis Pharma AG, Basel, Switzerland
5. Department of Electrical, Electronic and Information Engineering Guglielmo Marconi, University of Bologna, Bologna, Italy
6. Robert Bosch Gesellschaft für Medizinische Forschung, Stuttgart, Germany
7. Unit Digitale Geriatrie, Universitätsklinikum Heidelberg, Heidelberg, Germany
8. Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
9. Department of Neurology, University Medical Center Schleswig-Holstein Campus Kiel, Kiel, Germany
10. Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway
11. Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
12. The Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK

ARTICLE IN PRESS

ARTICLE IN PRESS

13. National Institute for Health and Care Research (NIHR) Newcastle Biomedical Research Centre (BRC), Newcastle University and The Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
14. Department of Physical Therapy, Gray Faculty of Medical & Health Sciences, Tel Aviv University, Tel Aviv, Israel
15. Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel
16. Rush Alzheimer's Disease Center and Department of Orthopaedic Surgery, Rush University Medical Center, Chicago, IL, USA

*Correspondence: Or Perlman, Email: orperlman@tauex.tau.ac.il

ARTICLE IN PRESS

ARTICLE IN PRESS

ABSTRACT

Physical activity and mobility are critical for healthy aging and predict diverse health outcomes. While wrist-worn accelerometers are widely used to monitor physical activity, estimating gait metrics from wrist data remains challenging. We extend ElderNet, a self-supervised deep-learning model previously validated for walking-bout detection, to estimate gait metrics from wrist accelerometry. Validation involved 819 older adults (Rush-Memory-and-Aging-Project) and 85 individuals with gait impairments (Mobilise-D), from six medical centers. In Mobilise-D, ElderNet achieved an absolute error of 8.82 cm/s and an intra-class correlation of 0.87 for gait speed, outperforming state-of-the-art methods ($p < 0.001$) and models using a lower-back sensor. ElderNet outperformed (percentage error; $p < 0.01$) competing approaches in estimating cadence and stride length, and better ($p < 0.01$) classified mobility disability (AUC = 0.80) than conventional gait or physical activity metrics. These results render ElderNet a scalable tool for gait assessment using wrist-worn devices in aging and clinical populations.

ARTICLE IN PRESS

ARTICLE IN PRESS

INTRODUCTION

Physical activity (PA) and mobility, the ability to move independently[1], are essential components of healthy aging and are strong predictors of a wide range of clinical outcomes, including cognitive decline, disability, and mortality[2–7]. Objective measures of PA—such as sedentary time and moderate-to-vigorous PA— are conventionally measured using wrist-worn accelerometers and offer valuable insights[8,9]. However, these measures typically capture overall physical activity levels and fail to distinguish between specific types of physical activity, such as gait versus other daily movements, limiting the ability to identify the distinct health benefits associated with different forms of activity[10].

Gait-specific measures, such as gait speed, cadence, and stride length, have been shown to serve as potential biomarkers, potentially providing complementary insight and objective measures of aging, functional health, and disease progression[11–14]. Thus, untangling gait from total daily physical activity may enable more precise identification of activity types to optimize the prediction of health outcomes in older populations and inform targeted interventions.

Traditionally, the assessment of gait quality has been conducted under controlled laboratory or clinical conditions, often using short walking tests or instrumented walkways[15,16]. While these methods offer high accuracy, they only reflect brief snapshots of an individual's mobility and may miss important variations that occur during daily life[17–19]. Recent advancements in wearable sensor technology have opened new opportunities to monitor gait in free-living, daily-life conditions, enabling long-term, remote assessment in natural environments[20,21]. Most real-world gait studies to date have relied on accelerometers placed on the lower back, trunk, legs or feet[15,22,23]. From a signal processing perspective, these sensor locations benefit from proximity to the body's center of mass or ground contact. These sensor placements allow for accurate spatial-temporal gait estimates[15], but comfort and usability concerns may limit long-term adherence and scalability[24].

In contrast, wrist-worn accelerometers are widely adopted in both clinical and consumer settings due to their convenience and high user compliance[9,16,25]. Large-scale population studies, such as

ARTICLE IN PRESS

the UK Biobank, have successfully demonstrated the feasibility of using wrist-worn accelerometers to derive meaningful population-level physical activity metrics, including step counts[6,9,26–28]. For example, Doherty et al. developed open-source algorithms for processing wrist accelerometry data, facilitating foundational insights into physical activity patterns across more than 100,000 participants[6,29–32]. However, these efforts have predominantly focused on overall physical activity quantification rather than on detailed assessments of spatial-temporal measures of gait, highlighting a critical gap and untapped potential.

The wrist's distance from the body's center of mass and the complexity of its motion during walking pose significant challenges for extracting meaningful and accurate gait metrics[33]. As a result, only a limited number of studies have attempted to estimate the spatial-temporal measures of gait from wrist data in real-world conditions. The few that do have largely focused on healthy individuals, rather than older adults or those with gait impairment[16,34], and often necessitate individual-level calibration[35,36].

Multi-day wrist recordings capture all of the movements of arm and hand. These recordings can include varied activities of the arm as well as walking. Two steps are necessary to extract meaningful gait metrics from wrist data. First, tools are needed to differentiate bouts of gait from non-gait movements throughout the entire recording. Once the bouts of gait are identified, additional tools are needed to quantify conventional gait characteristics, including speed, cadence, stride length, and regularity. In a recent study, we introduced and validated a self-supervised deep-learning framework, ElderNet, that accurately identifies bouts of gait from 3D wrist data[21]. ElderNet leverages a self-supervised learning (SSL) approach, utilizing large amounts of unlabeled accelerometry data to learn robust, generalizable representations of gait patterns. The framework comprises two main steps: an initial SSL-based pretraining using the large unlabeled UK Biobank dataset[6,9,26] (Dataset 1) followed by additional SSL training on the RUSH Memory and Aging Project (MAP) dataset[37,38] (Dataset 2) tailored toward older populations with diverse mobility characteristics and impairments. The second step involves fine-tuning the foundational SSL model for a specific downstream task, such as gait speed estimation, using a smaller, labeled dataset. Here, we adapted and expanded ElderNet to estimate detailed gait metrics for each identified walking bout, building on its previously demonstrated ability to detect walking segments from wrist-worn accelerometry.

ARTICLE IN PRESS

More specifically, we fine-tuned ElderNet using a labeled dataset from the Mobilise-D technical validation study (TVS; Dataset 3), a relatively large, daily living, labeled dataset comprising approximately 2.5-hours recordings from older adults and individuals with gait impairments, alongside robust reference annotations obtained from a validated multi-sensor reference system that included pressure insoles[39]. This dataset provides extensive real-world annotations of spatial-temporal gait parameters, enabling accurate model training and validation[40,41]. ElderNet's performance was evaluated on the test set of the TVS dataset (Dataset 3) against a fully supervised baseline model with the same architecture as ElderNet (but without SSL pretraining) to quantify the benefit of the SSL component. For gait speed estimation, we also compared ElderNet to a state-of-the-art biomechanical approach proposed by Soltani et al.[34], adapted to our dataset. In addition, to further assess generalizability, we validated ElderNet on an independent dataset of 11 healthy young adults (Dataset 4). Finally, to demonstrate ElderNet's potential clinical utility in real-world contexts, we applied it to a multi-day, free-living data from the RUSH MAP[37,38] (Dataset 2). Specifically, we examined whether the multi-day gait quality metrics extracted by ElderNet could effectively classify mobility disability status in older adults. Classifying mobility disability status in older adults using real-world gait metrics is essential for identifying individuals at elevated risk of motor and cognitive decline, gait-related falls, and other adverse outcomes[42,43]. We also directly compared the discriminative performance of multi-day gait metrics derived from ElderNet to metrics obtained from traditional supervised short assessments of gait and mobility, thereby demonstrating the added value of continuous real-world monitoring. This highlights differences between gait capacity (controlled assessments) and performance (daily behaviours), informing studies on real-world mobility disparities[18,44]. Furthermore, we compared the model based on multi-day gait metrics with a model using total daily physical activity metrics extracted from the same recordings to explore the added value of detailed multi-day gait metrics.

ARTICLE IN PRESS

ARTICLE IN PRESS

RESULTS

Evaluation of ElderNet for Gait Speed Estimation

To evaluate ElderNet's ability to estimate real-world gait quality metrics, we fine-tuned the model using the labeled dataset from the Mobilise-D TVS (Dataset 3). Figure 1 illustrates the ElderNet pipeline, and Table 1 summarizes the key characteristics of the Mobilise-D TVS dataset. Further details, including dataset composition and training protocols, are provided in Table 2 and the Methods section.

All reported results reflect performance on the Mobilise-D TVS test set (20% of dataset 3)**,** which includes representative individuals from each clinical cohort. On this test set, ElderNet demonstrated high performance in gait speed estimation, achieving median values across subjects for mean absolute error (MAE) of 8.8 cm/s, root mean square error (RMSE) of 10.7 cm/s, mean absolute percentage error (MAPE) of 12.73%, coefficient of determination (R^2) of 0.74, and, compared to measures obtained from the gold-standard pressure sensitive insole system, an intra-class correlation coefficient ICC(2,1) of 0.87. As shown in Figure 2, ElderNet significantly outperformed both the current state-of-the-art model by Soltani et al.[34] and a fully supervised baseline model, which uses the same architecture as ElderNet but is trained from scratch in a purely supervised manner, without any self-supervised pretraining phase. Furthermore, ElderNet maintained its superior performance on an external dataset (Dataset 4) comprising healthy young adults, achieving an MAE of 8.3 cm/s, RMSE of 11.4 cm/s, MAPE of 8.22%, R^2 of 0.72, and ICC of 0.87, and significantly outperforming the supervised baseline in both RMSE ($p<0.05$) and ICC ($p<0.01$). Detailed performance metrics, including the 25th and 75th percentiles at the subject level, are provided in Table S1. Figure 3 further illustrates these findings, showing a stronger correlation between ElderNet's predictions and true gait speeds, compared to alternative models.

ARTICLE IN PRESS

The Bland-Altman plot in Figure 4 compares ElderNet's gait speed predictions with reference measurements from various Mobilise-D TVS cohorts. This analysis revealed a minimal mean error of 0.17 cm/s, indicating low prediction bias, notably lower than a lower-back-based algorithm tested on the same cohorts[24]. Table 3 provides a summary of ElderNet's gait speed estimation performance across the different cohorts in the Mobilise-D TVS dataset. Post-hoc analysis showed stable gait speed errors across bouts of different durations (10–30, 30–60, 60–120, and >120 s; Kruskal-Wallis $p = 0.30$; Figure S1).

Sensitivity analysis

To assess the system's performance in a real-world deployment scenario where gait bouts are not pre-identified, we conducted an end-to-end validation on the Mobilise-D TVS (Internal Dataset 3) and an external validation set (Dataset 4). We applied our previously validated gait detection algorithm[21] on the continuous raw acceleration data to automatically identify gait bouts, which were then processed by the gait speed estimation model.

The subject-level mean gait speed derived from this fully automated pipeline showed strong agreement with the ground truth across both datasets (Dataset 3: MAE = 10.59 cm/s, ICC = 0.92; Dataset 4: MAE = 6.43 cm/s, ICC = 0.77) (Figure S2 A-B). Furthermore, to isolate the specific impact of detection errors, we compared the fully automated pipeline directly against the 'ideal' scenario (using ground-truth pre-identified walking bouts). This comparison revealed a near-perfect correlation (Dataset 3: Pearson's $r = 0.97$; Dataset 4: $r = 0.94$; $p < 0.001$ for both datasets) (Figure S2 C-D), confirming that errors propagated from the automated gait detection stage have a negligible impact on the final downstream gait speed estimates.

Evaluation of ElderNet for additional Gait Metrics

Beyond gait speed estimation, ElderNet was extended to quantify additional gait metrics, including cadence, stride length, and stride regularity, to provide a more comprehensive assessment of gait quality during daily living (Table 4). For cadence, ElderNet achieved a median MAE of 3.11 steps/min, RMSE of 5.19 steps/min, MAPE of 3.71%, R^2 of 0.65, and an ICC of 0.83. Compared with a fully supervised model, ElderNet showed significantly lower

ARTICLE IN PRESS

errors (MAE: $p < 0.001$; RMSE: $p = 0.045$), higher ICC ($p = 0.02$), and a trend towards a higher R^2, though this did not reach statistical significance ($p = 0.07$).

For stride length, ElderNet also demonstrated superior performance (MAE: 11.61 cm, $p < 0.01$; RMSE: 14.09 cm, MAPE: 11.09%, R^2: 0.48, ICC: 0.72; $p < 0.001$ for all others) compared with the supervised approach. Stride regularity estimation by ElderNet exhibited performance comparable to the supervised model, with both models achieving similar error rates (MAE = 0.08 A.U., RMSE = 0.11 A.U.; A.U. = arbitrary unit), though the supervised model indicated a slightly higher ICC (0.84 vs. 0.76) and R^2 (0.67 vs. 0.55). However, none of these differences were statistically significant (all p-values > 0.05).

Demonstration of Potential Clinical Utility

As a preliminary demonstration of ElderNet's potential clinical utility, we applied it to continuous, ~10-day wrist accelerometer recordings from the RUSH MAP dataset (Dataset 2), extracting daily-living gait quality measures (see Methods). Participant characteristics are summarized in Table 5. A heat map summarizing the correlations between measures obtained using the two approaches (Figure S3) revealed that gait characteristics recorded in supervised assessments are only mildly related to the same gait characteristics extracted from daily living conditions. Although internal consistency within settings was strong (supervised lower-back features correlated well among themselves, and similarly for wrist-based daily-living features), cross-setting correlations were weaker, suggesting that supervised gait measures do not fully reflect daily-living gait behavior.

To evaluate the discriminative utility of gait metrics collected in daily life, we compared performance against gait metrics obtained in a controlled, supervised setting, as well as PA features derived from daily living wrist accelerometry. We trained an XGBoost classifier to classify mobility disability status (defined using the Rosow-Breslau scale). As illustrated in Figure 5, the supervised-based model achieved an area under the curve (AUC) of 0.67 ± 0.09. In comparison, the model using PA features from daily-living wrist accelerometry performed slightly better (AUC: 0.69 ± 0.06). The highest accuracy was achieved by a model incorporating the daily life gait metrics (AUC: 0.80 ± 0.06).

ARTICLE IN PRESS

To statistically assess performance differences, we treated each outer cross-validation fold as an independent sample. A Friedman test revealed a significant overall effect of feature set on AUC ($p < 0.001$). Post hoc Wilcoxon signed-rank tests with a Bonferroni correction showed that the full daily-living gait model significantly outperformed the supervised assessment of gait model ($p < 0.01$) and also outperformed the PA model ($p < 0.01$). No significant difference was found between the supervised and PA models.

ARTICLE IN PRESS

ARTICLE IN PRESS

DISCUSSION

The aim of this study was to extend and validate ElderNet for extracting spatial-temporal measures of gait from a wrist-worn accelerometer during daily living in older adults and individuals with gait impairments. Prior studies have reported that gait speed metrics derived from a wrist-worn accelerometer are associated with incident frailty and dementia[13,45], yet the gait speed algorithms employed in prior studies have not previously been validated for use in older populations or those with gait impairments. Using real-world, daily-living wrist accelerometer recordings from a large cohort of well-characterized older adults and individuals with impaired gait, we optimized ElderNet to better support assessments in these groups. In addition, the current study includes preliminary clinical findings which demonstrate that multi-day mobility metrics extracted from a wrist-worn accelerometer can better discriminate adults with and without self-reported mobility disability. These metrics may improve subtyping of older adults compared to conventional gait performances and quantitative measures of total daily physical activity. Further work is needed to determine if different combinations of these metrics are differentially associated with disparate adverse health outcomes and can inform targeted therapies for impaired gait in aging adults.

For gait speed, ElderNet significantly outperformed the Soltani et al. state-of-the-art model[34], and a supervised baseline model by achieving lower error rates alongside high R^2 and ICC on the Mobilise-D TVS test set. Moreover, it achieved superior performance on an external dataset of healthy young adults. The improved performance by ElderNet was most pronounced within the Mobilise-D TVS cohort, highlighting the benefits of optimizing ElderNet models for assessing older adults and populations with gait impairments. Additionally, ElderNet consistently demonstrated high reliability across all cohorts, with ICC(2,1) values exceeding 0.7— a threshold commonly used in state-of-the-art Mobilise-D studies[39]. ElderNet's error remained stable across bout durations (recall Figure S1), affirming its robustness for real-world gait assessment.

In terms of clinical meaningfulness, the Minimal Clinically Important Difference (MCID) represents the smallest change in a given measure that is considered clinically significant, indicating a

ARTICLE IN PRESS

perceptible improvement or decline in a patient's condition or the effectiveness of an intervention. Prior research shows that the MCID for gait speed in adults with diverse health conditions, such as multiple sclerosis, acute cardiovascular disease, and hip fracture, typically falls within the range of 10 to 20 cm/s [46,47]. ElderNet's mean absolute error (MAE) of 8.8 cm/s (with RMSE of 10.7 cm/s and MAPE of 12.73%) falls below this threshold, suggesting that its estimation accuracy holds promise for detecting clinically meaningful changes in gait speed. However, given that individual-level monitoring is also influenced by day-to-day variability, future longitudinal studies are needed to validate the model's sensitivity to change over time. Moreover, although MAE falls below typical MCID thresholds, the Bland–Altman limits of agreement indicate that individual-level change detection requires dedicated longitudinal validation.

Fixed-placement 3D sensors on the lower back typically yield cleaner gait signals for multiday recordings compared to wrist-worn devices [20]. However, ElderNet has been shown to significantly reduce this performance gap[21]. Indeed, ElderNet also demonstrated improved accuracy compared to a lower-back-based gait speed algorithm previously evaluated on the Mobilise-D cohort by Kirk et al.[24]. Furthermore, ElderNet achieved a lower mean error of 0.17 cm/s, compared to 6.00 cm/s reported in the prior study using the lower back sensor. It is important to note that Kirk et al. evaluated their algorithm using all available data without splitting into separate training and test sets, complicating a direct comparison with our approach, which was strictly evaluated on an independent test set. Furthermore, Kirk et al. only included walking bouts that were successfully detected by both their gait detection algorithm and the reference system. This selection criterion likely favored simpler and less variable walking bouts, potentially leading to overly optimistic performance. In contrast, our validation included all walking bouts identified by the reference system, likely encompassing a broader range of complex and challenging gait scenarios. Thus, despite methodological differences, our approach may provide a more robust and ecologically valid estimation of real-world gait performance.

Gait speed estimation performance of ElderNet remained robust across all cohorts within the Mobilise-D dataset, but was lowest in the MS cohort, which was much younger than the other Mobilise-D participants. Reduced performance in the younger MS cohort (Table 1) may be due to the fact that ElderNet was explicitly optimized for older adults during the SSL phase, having been pretrained on the RUSH MAP dataset, which is comprised exclusively of older individuals

ARTICLE IN PRESS

(Table 5). In the future, if more labeled data becomes available, it may be interesting to refine ElderNet to improve its performance in healthier and younger adults.

We further fine-tuned ElderNet to estimate additional gait quality metrics—cadence, stride length, and stride regularity—capturing diverse facets of mobility[44,48,49]. To evaluate the contribution of the SSL component, we compared ElderNet's performance with that of a purely supervised baseline. For cadence and stride length, ElderNet consistently outperformed the supervised model across most evaluation metrics, with statistically significant improvements observed for all metrics except cadence R^2, which showed a positive trend. In contrast, no significant performance difference was observed for stride regularity. This may be due to the nature of the stride regularity labels, which were derived from lower-back accelerometer data, whereas the SSL phase of ElderNet relied solely on wrist-worn accelerometers. Incorporating multi-sensor data (e.g., gyroscopes, PPG[50]), including from additional body locations, into the SSL pre-training phase may enhance performance for such metrics. However, these enhancements would need to be balanced against increased system complexity, deployment costs, and the potential for reduced user compliance, and therefore warrant careful consideration in future research. In the meantime, the present findings suggest that representation learning enables inference of trunk-level gait dynamics from peripheral movement signals. This supports a broader paradigm in wearable sensing: distal sensors, when combined with large-scale self-supervised learning, may serve as scalable proxies for more proximal biomechanical measurements.

A potential advantage of deriving gait metrics from multi-day recordings is their ability to capture natural, within-subject variability in walking behavior, offering richer and more ecologically valid insights compared to traditional, short-duration supervised gait assessments. Here, we applied ElderNet on real-world data from the RUSH MAP, extracting meaningful statistics describing the distributions of the different gait quality metrics. Although gait metrics derived from supervised assessments were significantly associated with most of those obtained from daily living contexts, the observed correlations were weak-to-moderate (Figure S3). This confirms previous findings reported using lower back-worn sensors that supervised and real-world gait measures, while related, likely reflect distinct aspects of gait and mobility[17–19]. The

ARTICLE IN PRESS

ability to extract accurate gait measures from a wrist-worn accelerometer in daily life, therefore, holds promise for extending these insights and improving diagnostic and prognostic capabilities.

Indeed, our analysis demonstrated that gait metrics derived from multi-day recordings provided superior discriminative power for mobility disability compared to conventional supervised assessments. Moreover, these real-world gait metrics also outperformed the model based on PA metrics derived from the wrist-worn accelerometer from which the gait metrics were derived. This highlights the added clinical value of extracting gait-specific measures from everyday living metrics, offering a more detailed and predictive assessment of mobility and health status.

It is worth noting that the clinical utility analyses in the current study were based on a basic machine learning approach, serving as a preliminary demonstration of clinical potential. Future research will aim to develop dedicated neural network-based classifiers specifically trained on clinical outcomes, integrating and potentially fine-tuning ElderNet directly on tasks such as mobility disability classification or the prediction of the development of adverse events such as mortality or the development of Alzheimer's or Parkinson's using multiday recordings of daily living. We anticipate that the application of more advanced deep learning techniques will further leverage the benefits of extracting gait measures from a wrist-worn sensor.

One limitation of this study is that ElderNet was validated on a test set drawn from the same distribution as the training dataset (i.e., the Mobilise-D dataset). In addition, the test set comprised a relatively small sample size (n=18), which may affect the robustness of our conclusions. Furthermore, the UK Biobank and Mobilise-D cohorts primarily represent Western populations, which may limit the model's generalizability to underrepresented racial, ethnic, or socioeconomic groups. To partially address these limitations, we employed an external dataset of healthy young adults (n=11) and further utilized the RUSH MAP dataset (n=423) to extend the validation. Nonetheless, future studies should further investigate ElderNet's effectiveness on additional external labeled, diverse datasets and, if needed, aim to further enhance its generalizability and clinical utility.

We also note that while ElderNet's self-supervised pretraining incorporated data from diverse accelerometer devices (Datasets 1 and 2), and training included augmentations to encourage invariance to sensor orientations, the primary fine-tuning and validation were conducted on a

ARTICLE IN PRESS

single device type (Datasets 3 and 4). Table S2 provides detailed device specifications for the different datasets used in this study. ElderNet's design leverages multi-device exposure, with minimal, common device specifications, in the foundational SSL phase to potentially improve cross-device robustness for real-world deployment. However, future work should explicitly evaluate performance across a broader range of commercial wrist-worn devices and sensor specifications to confirm this generalizability.

Recent work with UK Biobank accelerometry data highlights the potential feasibility of using daily living information to stratify older adults at risk for Parkinson's disease[51,52]. However, this study did not exploit detailed multi-day gait-specific measures and suffered from other limitations. By integrating spatial-temporal gait metrics derived from multi-day recordings, there is an opportunity to significantly enhance the early detection and characterization of such conditions.

Overall, our findings demonstrate that ElderNet, previously validated for accurately detecting walking bouts from wrist-worn accelerometers, can now reliably estimate spatial-temporal gait metrics during everyday life. This capability has been specifically validated in older adults with age-related gait impairments and neurological conditions. Utilizing wrist-worn accelerometers for remote phenotyping offers substantial practical advantages, including affordability, widespread acceptance by older adults, high compliance, and suitability for long-term unobtrusive monitoring. Historically, detailed gait characterization has been challenging due to the complexity of wrist-worn accelerometer signals—a limitation directly addressed by our approach. By overcoming these barriers, ElderNet advances aging research, enabling the identification of specific gait alterations that may serve as biomarkers for diseases or functional impairments. Ultimately, this could facilitate the development of personalized interventions informed by an individual's unique mobility patterns.

ARTICLE IN PRESS

METHODS

Study Design

This study consisted of three primary stages:

1. **Estimation of Spatial-Temporal Gait Measures**: Extending ElderNet to quantify continuous gait quality metrics.
2. **Model Validation and Comparison**: Comparing ElderNet's performance against state-of-the-art and baseline supervised models using internal and external datasets.
3. **Demonstration of Potential Clinical Utility**: Evaluating the clinical relevance of gait metrics by applying ElderNet to real-world daily living data from cohorts with varying clinical characteristics.

Datasets and Participants

Four independent datasets were used across different phases of model development, validation, and evaluation of clinical utility:

Dataset 1: UK Biobank, a large-scale dataset comprising one week of wrist-worn accelerometry recordings (Axivity AX3, 100 Hz sampling rate) from over 100,000 individuals (mean age 62.1 ± 7.9 years). The UK Biobank dataset was used indirectly via the publicly available ResNet deep learning model pre-trained using SSL by Yuan et al.[26]. This model served as the fixed feature extraction backbone for ElderNet.

Dataset 2: Rush Memory and Aging Project (MAP), which includes 819 older adults (mean age 83.4 ± 7.3 years) who wore a wrist accelerometer (Axivity AX3, 50 Hz sampling rate or GENEActiv, 40 Hz sampling rate) for up to 10 consecutive days. This dataset was used in two stages: (i) to adapt ElderNet to the mobility patterns of older adults through additional SSL pretraining, and (ii) to evaluate the clinical utility of ElderNet by classifying mobility disability based on multi-day gait metrics. A subset of 423 individuals (mean age 83.6 ± 6.9 years) also underwent standardized supervised gait assessments, enabling comparison between real-world and clinical gait measures.

ARTICLE IN PRESS

Dataset 3: Mobilise-D Technical Validation Study. This multicenter observational study comprised 108 participants recruited from five clinical centers across the UK, Germany, and Israel. Following the exclusion of 23 individuals due to incomplete recordings, 85 participants remained for analysis. The cohort included both healthy older adults and individuals diagnosed with one of five conditions: chronic obstructive pulmonary disease (COPD), Parkinson's disease (PD), multiple sclerosis (MS), proximal femoral fracture (PFF), and congestive heart failure (CHF). Each participant wore an inertial unit including a tri-axial accelerometer on the non-dominant wrist, a lower-back inertial unit, and the INDIP multi-sensor reference system for up to 2.5 hours during their typical daily activities (at home, work, in the community, or outdoors). Participants were also encouraged to perform a limited set of predefined tasks (outdoor walking, walking up and down a slope and stairs, and moving from one room to another) [53]. The INDIP system, which integrates complementary sensing modalities—including two plantar pressure insoles, three inertial units, and two distance sensors—was used to detect walking bouts and provide ground-truth annotations of spatial-temporal gait metrics at the stride level[40,41].

This dataset was used for hyperparameter tuning, training, and evaluation of ElderNet. It was partitioned into training (70%), validation (10%), and test (20%, n=18) sets, with subject-wise partitioning to prevent data leakage. All three sets—training, validation, and test—include representation from all clinical cohorts in the Mobilise-D study (i.e., HA, CHF, COPD, MS, PD, and PFF), ensuring balanced evaluation across the diverse participant population. All performance comparisons—including against baseline supervised and state-of-the-art models—were conducted on the held-out test set. Ethical approvals for data collection were obtained from the relevant institutional review boards at all participating sites. A complete description of the data collection protocol is available in Mazza et al.[39].

Dataset 4: Mobilise-D Healthy Young Adults (External Validation). An independent validation dataset comprising 11 healthy young adults (mean age 29.6 ± 7.8 years) , collected using the same protocol as the Mobilise-D TVS. These data were used to assess ElderNet's generalizability in gait speed estimation to a younger, healthier, and demographically distinct population.

For the Rush Memory and Aging project, all participants provided written informed consent before participation. The study was approved by the Rush University Medical Center

ARTICLE IN PRESS

Institutional Review Board and conducted in accordance with the Declaration of Helsinki. For the Mobilise-D study, ethical approval was obtained at the individual sites (London-Bloomsbury Research Ethics Committee, 19/LO/1507; Helsinki Committee, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel, 0551-19TLV; ethical committee of the medical faculty of The University of Tübingen, 647/2019BO2; ethical committee of the medical faculty of Kiel University, D438/18) and all participants gave informed consent before participating.

Data Preprocessing

The INDIP reference system was first used to identify the exact start- and end-times of every walking bout within the approximately 2.5-hour recording[20]. The corresponding wrist-accelerometer segments were extracted, ensuring that non-walking periods were discarded. Each consecutive bout lasting longer than 10 seconds was then partitioned into 10-second windows with 90 % overlap (i.e., a stride of 1 second), where windows never spanned two different bouts. We chose 10-second windows to align with the pre-trained UK Biobank model[26] and added 90% overlap to enhance the temporal resolution of our analysis. Importantly, while overlapping windows were used for model training and visualization, all statistical comparisons were performed at the subject level after aggregating window-level predictions, ensuring that the correlated structure of overlapping windows does not affect inferential statistics. Although window-level samples are highly correlated due to overlap, model training was performed with subject-wise partitioning and cross-validation, preventing leakage of temporally adjacent windows across folds.

All extracted segments were resampled to 30 Hz, matching ElderNet's pre-training data from UK Biobank[26]. Because typical gait frequencies are below 10 Hz, this rate satisfies the Nyquist criterion and preserves all relevant signal content. To assess the impact of walking bout length on gait speed estimation, a post-hoc analysis was conducted in which gait speed error was stratified by bout duration (10–30 s, 30–60 s, 60–120 s, >120 s) in the Mobilise-D TVS test set (Dataset 3).

Wrist-worn devices varied across datasets, partially reflecting real-world diversity. Beyond resampling all signals to 30 Hz, we also performed a data-augmentation step during training. Each triaxial training sample was randomly rotated and its axes were switched, which promoted

ARTICLE IN PRESS

rotation invariance and enhanced the model's ability to generalize across devices with different sensor orientations.

Spatial-Temporal Gait Metrics

For every 10-second walking window, we derived four reference outcomes from the INDIP system: gait speed (cm/s), stride length (cm), cadence (steps/min), and stride regularity (arbitrary units, on a scale of 0 to 1). Speed and stride length were defined as the median of all complete INDIP-detected strides whose start and end both fell inside the window. Cadence was obtained by dividing the number of strides by the 10-second window duration and multiplying by six (one stride equals two steps). Stride regularity was calculated following Moe-Nilssen and Helbostad[54]. Briefly, the vertical acceleration of the lower-back device was demeaned and normalized, an unbiased autocorrelation function was applied to the 10-second segment, and the coefficient at a time lag corresponding to the mean stride duration (the second dominant peak) was taken as the stride-regularity index (range 0–1, with 1 denoting perfectly periodic strides). Each analysis window, therefore, carried a fully time-aligned ground-truth label for all four gait outcomes. For all analyses requiring bout-level metrics, the value for a specific bout was calculated as the median of the predicted values across all 10-second windows comprising that bout.

ElderNet Model Development and Training

ElderNet, originally developed as a classification model for gait detection[21], was adapted for the regression-based estimation of continuous gait metrics through a two-stage process. In the first stage, we leveraged a ResNet model that was pre-trained using SSL on the large-scale UK Biobank wrist-worn accelerometry dataset[6,26]. This pre-trained backbone served as a fixed feature extractor, with all ResNet weights frozen. To adapt the model for older adults, we added three fully connected layers on top of the frozen backbone and trained these layers using SSL in a multi-task learning framework on data from the RUSH MAP, which includes wrist-worn accelerometer recordings of older populations. The output of this stage was a 128-dimensional feature vector for each 10-second accelerometry window.

In the second stage, this RUSH-tailored SSL model was repurposed as a feature extractor for gait quality estimation. To enable regression outputs, we extended the model with 0–3 additional layers (selected via hyperparameter tuning) that mapped the 128-dimensional feature representation

ARTICLE IN PRESS

to continuous gait metrics such as gait speed. During this fine-tuning stage, the entire model, including the ResNet backbone and the newly added layers, was trained end-to-end using the training data from the Mobilise-D dataset (Dataset 3). This approach allowed ElderNet to learn precise, task-specific mappings while benefiting from the robust representations learned during the SSL phase.

Because stride length can also be expressed as walking speed divided by cadence, we computed two independent stride-length estimates for every 10-second window: one taken directly from the ElderNet output trained on stride-length labels, and a second obtained by dividing ElderNet's predicted speed by its predicted cadence. We then reported the simple average of these two values as the final stride-length prediction, an approach that consistently reduced error compared with relying on either estimate alone. We did not systematically evaluate alternative fusion strategies (e.g., weighted averaging or learned fusion), which could be explored in future work.

Hyperparameter optimization was performed on the training (70% of dataset 3) and validation (10% of dataset 3) sets using the Optuna Python library with 50 trials[55], employing five-fold cross-validation stratified by clinical cohort and grouped by participants. The hyperparameter search space is detailed in Supplementary Table S3. Model training incorporated early stopping with a patience of five epochs, selecting the configuration yielding the lowest MAE on the validation set. The loss function used was L1.

The model was implemented using PyTorch in Python 3.10 and trained using an NVIDIA RTX A4000 GPU with 16GB RAM. The training dataset consisted of 49,372 windows of 10 seconds each. During the final training with the optimal hyperparameter configuration, the average processing throughput was approximately 2,541.15 windows per second. For inference, the test set contained 13,998 windows, with a total inference time recorded at 8.55 seconds.

Model Evaluation and Comparison

ElderNet's performance was evaluated against a fully supervised baseline model with an identical ResNet architecture (without SSL pre-training) to quantify the benefit derived from the SSL component. Additionally, for gait speed, we compared ElderNet to the non-personalized biomechanically-based approach proposed by Soltani et al.[34]. This approach uses regression-

ARTICLE IN PRESS

based modeling with predefined biomechanical features extracted from wrist-worn accelerometry data, including cadence, acceleration energy, mean absolute jerk, hand-swing intensity, mean acceleration norm, and altitude change of path. Given the absence of barometric data in our dataset, altitude-related features were excluded from the model. We trained Soltani's regression model using the same training dataset as ElderNet (i.e., 80% of dataset 3) and evaluated it on the identical test dataset (i.e., 20% of dataset 3, and external dataset 4) to ensure a fair and consistent comparison. We considered the method proposed by Soltani et al. as a state-of-the-art benchmark due to its strong real-world performance in gait speed estimation, achieving a median RMSE of approximately 10 cm/s and demonstrating significant associations with clinical metrics such as frailty and handgrip strength[45].

Demonstration of Potential Clinical Utility

To evaluate the clinical utility of gait metrics derived from ElderNet, we applied the fine-tuned models to real-world wrist accelerometry data collected during daily life from 819 community-dwelling older adults participating in the RUSH MAP[37,38] (Dataset 2). Among these participants, 423 underwent additional supervised gait assessments using lower-back accelerometry during standardized 8-foot and 32-foot walking tests. For each of the supervised walking tests, we extracted four measures corresponding to the daily living gait quality metrics: gait speed, stride length, cadence, and stride regularity[56,57].

Our aim was to determine whether multi-day gait metrics could capture aspects of mobility that are not fully represented by brief supervised assessments or general physical activity measures. Daily-living walking bouts were identified from the wrist-worn accelerometry recordings using the previously validated gait detection algorithm implemented in ElderNet[21]. From each identified walking bout, we extracted four gait quality metrics—gait speed, stride length, cadence, and stride regularity—using the fine-tuned ElderNet regression models. Additionally, stride regularity was independently calculated directly from the raw wrist accelerometry data using an autocorrelation-based signal processing method analogous to the approach applied to lower-back data[54].

To accurately analyze participant activity, we excluded time segments during which participants were not wearing the device. These non-wear periods were defined as consistent low movement

ARTICLE IN PRESS

(low STD) across all acceleration axes for at least 30 minutes, as previously described[2,58]. Participants with fewer than three valid days of wrist accelerometry data were excluded, resulting in a final sample of 787 participants[9]. For each participant, comprehensive summary statistics—including mean, median, standard deviation, kurtosis, skewness, range, and decile percentiles—were calculated at both the 10-second window level and across aggregated walking bouts. In parallel, PA measures were derived from the magnitude of the same 3D acceleration signal, summarized over a daily time frame. These daily summaries included the same set of statistics used for gait quality metrics and were computed for both the entire recordings (to capture non-walking PA) and the segments identified as walking bouts. Specifically, the acceleration magnitude was calculated as the Euclidean vector magnitude of the triaxial signal. Vector magnitude serves as a standard proxy for overall PA intensity, capturing the net body movement independent of directionality, and has been widely validated for estimating energy expenditure and activity levels in wrist-worn accelerometry[6]. For non-gait PA, participant-level features were then aggregated across valid days (e.g., median and standard deviation of daily vector magnitude values), providing insights into day-to-day variability in overall activity that complements gait-specific metrics by reflecting broader functional mobility, such as engagement in upper-body or light non-locomotor tasks. A detailed overview of all extracted features is provided in Table S4.

Finally, to assess the clinical relevance of the derived gait metrics, we constructed classification models using an XGBoost classifier (implemented via the scikit-learn library) to discriminate between participants with and without mobility disability, as defined by the Rosow-Breslau scale (dichotomized into no disability versus disability)[42]. Three separate classification models were implemented: one based on gait features from supervised assessments, a second using physical activity features from daily living recordings, and a third incorporating daily living gait metrics. Model performance was evaluated using nested cross-validation with 5-fold inner and 10-fold outer loops with subject-wise partitioning to prevent data leakage. Hyperparameter tuning was performed via a randomized search (see Table S5 for the hyperparameter ranges). Model evaluation was based on Receiver Operating Characteristic (ROC) analysis, with the AUC serving as the primary metric.

Statistical Analysis

ARTICLE IN PRESS

Statistical analyses for the estimation of gait metrics (Datasets 3 and 4) were conducted at the subject level: model-performance metrics were first aggregated over the repeated 10-s windows contributed by each participant, so every statistical test compares independent subjects. Shapiro–Wilk tests rejected normality; therefore, non-parametric, two-sided tests were used throughout.

For gait speed, where three models (ElderNet, a supervised baseline, and a state-of-the-art comparator) were compared, we used the Friedman test to assess overall differences across models. If the Friedman test indicated statistical significance, post hoc pairwise comparisons were conducted using the two-sided Wilcoxon signed-rank test, with Bonferroni correction applied to account for multiple comparisons. For the other gait metrics (cadence, stride length, and stride regularity), comparisons involved only two models (ElderNet and its supervised counterpart). Therefore, we used the Wilcoxon signed-rank test to evaluate differences in subject-level performance, with p-values adjusted for multiple comparisons via the Bonferroni method.

For the classification task in the RUSH MAP (Dataset 2), the AUC from each of the ten outer cross-validation folds was treated as an independent observation. To compare the three competing feature sets (supervised gait metrics, PA metrics, and daily-living gait metrics), we employed the Friedman test on fold-level AUCs. When the Friedman test revealed significant differences, we conducted post hoc two-sided Wilcoxon signed-rank tests with Bonferroni correction for pairwise comparisons.

All statistical analyses were performed in Python using the scipy.stats library. Statistical significance was set at $p < 0.05$ after correction for multiple comparisons. ICC values were calculated to assess the agreement between ElderNet's predicted gait metrics and the reference measurements from the INDIP system. ICC values were interpreted based on established guidelines: ICC < 0.5 indicates poor reliability, 0.5–0.75 indicates moderate reliability, 0.75–0.9 indicates good reliability, and ICC > 0.9 indicates excellent reliability[59].

ARTICLE IN PRESS

Data availability

The dataset from the Mobilise-D technical validation study (Dataset 3) can be found on Zenodo: https://doi.org/10.5281/zenodo.13899386. All other data and related algorithms (i.e., Dataset 2) included in these analyses are available via the Rush Alzheimer's Disease Center Research Resource Sharing Hub, which can be found at www.radc.rush.edu. It has descriptions of the studies and available data. Any qualified investigator can create an account and submit requests for de-identified data.

Code availability

The underlying code for this study is available on Github and can be accessed via this link: https://github.com/yonbrand/gait-quality/tree/main

Acknowledgments

The authors gratefully acknowledge the entire Mobilise-D Work Package 2 team for their ongoing collaboration, insightful discussions, and valuable contributions. Special thanks are extended to the study participants for their time, commitment, and enthusiasm, especially during the challenging circumstances of the COVID-19 pandemic. The authors also sincerely thank the contributors to the RUSH Memory and Aging Project, as well as the dedicated staff at the Rush Alzheimer's Disease Center for their continuous support.

This work was supported in part by grants from the NIH (R01AG017917; R01AG056352, R01AG79133, R01AG078256), the Aufzien Family Center for the Prevention and Treatment of Parkinson's Disease (APPD) at Tel Aviv University, and the Mobilise-D project. The Mobilise-D project received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 820820. This JU receives support from the European Union's Horizon 2020 research and innovation program and the European Federation of Pharmaceutical Industries and Associations (EFPIA). SDD, AJY, and LR were supported by the IDEA-FAST project that has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 853981. SDD, AJY, and LR were supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at The Newcastle

ARTICLE IN PRESS

upon Tyne Hospital NHS Foundation Trust, Newcastle University, and the Cumbria, Northumberland and Tyne and Wear (CNTW) NHS Foundation Trust. SDD, AJY, and LR were also supported by the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon Tyne Hospitals NHS Foundation Trust. SDD was supported by the UK Research and Innovation (UKRI) Engineering and Physical Sciences Research Council (EPSRC) (Grant Ref: EP/X031012/1 and Grant Ref: EP/X036146/1).

All opinions are those of the authors and not the funders. The content in this publication reflects the authors' view, and neither IMI nor the European Union, EFPIA, NHS, NIHR, or any associated partners are responsible for any use that may be made of the information contained herein.

Author Contributions

Participant recruitment and clinical oversight: W.M., C.B, J.M.H., A.J.Y., L.R. Algorithm development: Y.E.B., O.P., J.M.H. Data analysis, statistical analysis: Y.E.B. Figures and tables preparation: Y.E.B. Data interpretation: Y.E.B., J.M.H., O.P. Drafting of the initial manuscript: Y.E.B., J.M.H., O.P. Intellectual contribution: Y.E.B., F.K., L.P., C.B., A.C., W.M., B.V., A.J.Y., L.R., S.DD., A.M., A.S.B., J.M.H., O.P. All authors have provided critical intellectual input during the revision of the manuscript. All authors have reviewed the manuscript and approved the submitted version.

Competing interests

S. Del Din reports consultancy activity with Hoffmann-La Roche Ltd. outside of this study.

ARTICLE IN PRESS

REFERENCES

1. Reijnierse, E. M. *et al.* Towards a core-set of mobility measures in ageing research: The need to define mobility and its constructs. *BMC Geriatr* 23, 220 (2023).
2. Sabia, S. *et al.* Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors. *Am J Epidemiol* 179, 781–790 (2014).
3. Buchman, A. S. *et al.* Physical activity, common brain pathologies, and cognition in community-dwelling older adults. *Neurology* 92, E811–E822 (2019).
4. Zou, Q. *et al.* Longitudinal Association between Physical Activity, Blood Lipids, and Risk of Dyslipidemia among Chinese Adults: Findings from the China Health and Nutrition Surveys in 2009 and 2015. *Nutrients* 15, 341 (2023).
5. Buchman, A. S. Total Daily Physical Activity and Longevity in Old Age. *Arch Intern Med* 172, 444-450 (2012).
6. Doherty, A. *et al.* Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. *PLoS One* 12, e0169649 (2017).
7. Zhao, A., Cui, E., Leroux, A., Lindquist, M. A. & Crainiceanu, C. M. Evaluating the prediction performance of objective physical activity measures for incident Parkinson's disease in the UK Biobank. *J Neurol* 270, 5913–5923 (2023).
8. Leroux, A. *et al.* Organizing and analyzing the activity data in NHANES. *Stat Biosci* 11, 262–287 (2019).
9. Willetts, M., Hollowell, S., Aslett, L., Holmes, C. & Doherty, A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. *Scientific Reports* 8, 7961 (2018).
10. Lin, W. *et al.* Can Gait Characteristics Be Represented by Physical Activity Measured with Wrist-Worn Accelerometers? *Sensors* 23, 8542 (2023).
11. Studenski, S. Gait Speed and Survival in Older Adults. *JAMA* 305, 50-58 (2011).
12. Dumurgier, J. *et al.* Gait Speed and Decline in Gait Speed as Predictors of Incident Dementia. *The Journals of Gerontology: Series A* 72, 655–661 (2017).
13. Chan, L. L. Y., Espinoza Cerda, M. T., Brodie, M. A., Lord, S. R. & Taylor, M. E. Daily-life walking speed, running duration and bedtime from wrist-worn sensors predict incident dementia: A watch walk – UK biobank study. *Int Psychogeriatr* 100031 (2025) doi:10.1016/J.INPSYC.2024.100031.
14. Weiss, A. *et al.* Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. *Neurorehabil Neural Repair* 25, 810–818 (2011).
15. Zadka, A. *et al.* A wearable sensor and machine learning estimate step length in older adults and patients with neurological disorders. *npj Digital Medicine* 7, 142 (2024).
16. Chan, L. L. Y., Choi, T. C. M., Lord, S. R. & Brodie, M. A. Development and large-scale validation of the Watch Walk wrist-worn digital gait biomarkers. *Scientific Reports* 12, 16211 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

17. Hillel, I. *et al.* Is every-day walking in older adults more analogous to dual-task walking or to usual walking? Elucidating the gaps between gait performance in the lab and during 24/7 monitoring. *Eur Rev Aging Phys Act* 16, 6 (2019).

18. Warmerdam, E. *et al.* Long-term unsupervised mobility assessment in movement disorders. *Lancet Neurol* 19, 462–470 (2020).

19. Hausdorff, J. M. *et al.* Everyday Stepping Quantity and Quality Among Older Adult Fallers With and Without Mild Cognitive Impairment: Initial Evidence for New Motor Markers of Cognitive Deficits? *J Gerontol A Biol Sci Med Sci* 73, 1078–1082 (2018).

20. Kluge, F. *et al.* Real-World Gait Detection Using a Wrist-Worn Inertial Sensor: Validation Study. *JMIR Form Res 2024;8:e50035 https://formative.jmir.org/2024/1/e50035* 8, e50035 (2024).

21. Brand, Y. E. *et al.* Self-supervised learning of wrist-worn daily living accelerometer data improves the automated detection of gait in older adults. *Scientific Reports* 14, 20584 (2024).

22. Bonci, T., Keogh, A., Din, S. Del, Scott, K. & Mazzà, C. An Objective Methodology for the Selection of a Device for Continuous Mobility Assessment. *Sensors* 20, 6509 (2020).

23. Weston, A. R. *et al.* Turning speed as a more responsive metric of age-related decline in mobility: A comparative study with gait speed. *Clinical Biomechanics* 113, 106196 (2024).

24. Kirk, C. *et al.* Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. *Scientific Reports* 14, 1754 (2024).

25. Adams, J. L. *et al.* Using a smartwatch and smartphone to assess early Parkinson's disease in the WATCH-PD study over 12 months. *npj Parkinson's Disease* 10, 112 (2024).

26. Yuan, H. *et al.* Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *npj Digital Medicine* 7, 91 (2024).

27. Acquah, A. *et al.* Daily steps are a predictor of, but perhaps not a modifiable risk factor for Parkinson's Disease: findings from the UK Biobank. *medRxiv* 2024.08.13.24311539 (2024) doi:10.1101/2024.08.13.24311539.

28. Small, S. R. *et al.* Development and Validation of a Machine Learning Wrist-worn Step Detection Algorithm with Deployment in the UK Biobank. *medRxiv* 2023.02.20.23285750 (2023) doi:10.1101/2023.02.20.23285750.

29. Shreves, A. H. *et al.* Amount and intensity of daily total physical activity, step count and risk of incident cancer in the UK Biobank. *Br J Sports Med* 59, 839-847 (2025) doi:10.1136/BJSPORTS-2024-109360.

30. Barker, J. *et al.* Physical activity of UK adults with chronic disease: cross-sectional analysis of accelerometer-measured physical activity in 96 706 UK Biobank participants. *Int J Epidemiol* 48, 1167–1174 (2019).

31. Stamatakis, E. *et al.* Association of wearable device-measured vigorous intermittent lifestyle physical activity with mortality. *Nat Med* 28, 2521–2529 (2022).

ARTICLE IN PRESS

32. Bloomberg, M., Brocklebank, L., Doherty, A., Hamer, M. & Steptoe, A. Associations of accelerometer-measured physical activity, sedentary behaviour, and sleep with next-day cognitive performance in older adults: a micro-longitudinal study. *International Journal of Behavioral Nutrition and Physical Activity* 21, 133 (2024).

33. Brand, Y. E. *et al.* Gait Detection from a Wrist-Worn Sensor Using Machine Learning Methods: A Daily Living Study in Older Adults and People with Parkinson's Disease. *Sensors* 22, 7094 (2022).

34. Soltani, A., Dejnabadi, H., Savary, M. & Aminian, K. Real-World Gait Speed Estimation Using Wrist Sensor: A Personalized Approach. *IEEE J Biomed Health Inform* 24, 658–668 (2020).

35. Urbanek, J. K. *et al.* Free-Living Gait Cadence Measured by Wearable Accelerometer: A Promising Alternative to Traditional Measures of Mobility for Assessing Fall Risk. *The Journals of Gerontology: Series A* 78, 802–810 (2023).

36. Karas, M. *et al.* Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation. *Biostatistics* 22, 331–347 (2021).

37. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's Disease* 64, S161–S189 (2018).

38. Bennett, D.A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr Alzheimer Res* 9, 646–663 (2012).

39. Mazzà, C. *et al.* Technical validation of real-world monitoring of gait: a multicentric observational study. *BMJ Open* 11, e050785 (2021).

40. Salis, F. *et al.* A method for gait events detection based on low spatial resolution pressure insoles data. *J Biomech* 127, 110687 (2021).

41. Salis, F. *et al.* A multi-sensor wearable system for the assessment of diseased gait in real-world conditions. *Front Bioeng Biotechnol* 11, 1143248 (2023).

42. Buchman, A. S., Boyle, P. A., Leurgans, S. E., Evans, D. A. & Bennett, D. A. Pulmonary function, muscle strength, and incident mobility disability in elders. *Proc Am Thorac Soc* 6, 581–587 (2009).

43. Buchman, A. S., Boyle, P. A., Leurgans, S. E., Barnes, L. L. & Bennett, D. A. Cognitive function is associated with the development of mobility impairments in community-dwelling elders. *American Journal of Geriatric Psychiatry* 19, 571–580 (2011).

44. Shema-Shiratzky, S. *et al.* A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity. *J Neurol* 267, 1912–1921 (2020).

45. Guralnik, J. M., Ferrucci, L., Simonsick, E. M., Salive, M. E. & Wallace, R. B. Lower-Extremity Function in Persons over the Age of 70 Years as a Predictor of Subsequent Disability. *New England Journal of Medicine* 332, 556–561 (1995).

46. Soltani, A. *et al.* Real-world gait speed estimation, frailty and handgrip strength: a cohort-based study. *Scientific Reports* 11, 18966 (2021).

ARTICLE IN PRESS

ARTICLE IN PRESS

47. Tamura, S. *et al.* Minimal clinically important difference of the short physical performance battery and comfortable walking speed in old-old adults with acute cardiovascular disease: a multicenter, prospective, observational study. *Disabil Rehabil* 45, 1079–1086 (2023).

48. Bohannon, R. W. & Glenney, S. S. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. *J Eval Clin Pract* 20, 295–300 (2014).

49. Verghese, J., Holtzer, R., Lipton, R. B. & Wang, C. Quantitative Gait Markers and Incident Fall Risk in Older Adults. *The Journals of Gerontology: Series A* 64A, 896–901 (2009).

50. Lord, S., Galna, B. & Rochester, L. Moving forward on gait measurement: Toward a more refined approach. *Movement Disorders* 28, 1534–1543 (2013).

51. Asher, E. E. *et al.* Combining 24-Hour Continuous Monitoring of Time-Locked Heart Rate, Physical Activity and Gait in Older Adults: Preliminary Findings. *Sensors* 25, 1945 (2025).

52. Schalkamp, A. K., Peall, K. J., Harrison, N. A. & Sandor, C. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nature Medicine* 29, 2048–2056 (2023).

53. Mirelman, A., Rochester, L., Simuni, T. & Hausdoff, J. M. Digital mobility measures to predict Parkinson's disease. *Lancet Neurol* 22, 1098–1100 (2023).

54. Scott, K. *et al.* Design and validation of a multi-task, multi-context protocol for real-world gait simulation. *J Neuroeng Rehabil* 19, 141 (2022).

55. Moe-Nilssen, R. & Helbostad, J. L. Estimation of gait cycle characteristics by trunk accelerometry. *J Biomech* 37, 121–126 (2004).

56. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2623–2631 (2019) doi:10.1145/3292500.3330701.

57. Hill, E. J. *et al.* Quantitative mobility measures complement the MDS-UPDRS for characterization of Parkinson's disease heterogeneity. *Parkinsonism Relat Disord* 84, 105–111 (2021).

58. Buchman, A. S. *et al.* Associations between Quantitative Mobility Measures Derived from Components of Conventional Mobility Testing and Parkinsonian Gait in Older Adults. *PLoS One* 9, e86262 (2014).

59. van Hees, V. T. *et al.* Estimation of Daily Energy Expenditure in Pregnant and Non-Pregnant Women Using a Wrist-Worn Tri-Axial Accelerometer. *PLoS One* 6, e22922 (2011).

60. Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 15, 155–163 (2016).

61. File:Circle-icons-speedometer.svg - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Circle-icons-speedometer.svg

62. File:Dall-e 3 (jan '24) artificial intelligence icon.png - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Dall-e_3_(jan_%2724)_artificial_intelligence_icon.png

ARTICLE IN PRESS

ARTICLE IN PRESS

63. File:Silhouette of man with cane.svg - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Silhouette_of_man_with_cane.svg

64. File:Homoj.svg - Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Homoj.svg

Figure 1. Overview of the ElderNet pipeline for gait metric estimation from wrist-worn accelerometer data. The pipeline begins with self-supervised learning pretraining on the UK Biobank dataset, followed by adaptation for older adults using data from the RUSH Memory and Aging Project. The adapted ElderNet model is then fine-tuned for gait quality estimation using labeled data from the Mobilise-D dataset, incorporating a regression head to derive gait-related features from acceleration signals. Incorporates elements from Wikimedia Commons under open licenses[60–63].

Figure 2: Gait speed prediction performance comparison of ElderNet with benchmark models. Boxplots show subject-level error and performance metrics for ElderNet (green boxes and circles), state-of-the-art (orange boxes and circles), and supervised baseline (blue boxes and circles) models on internal Mobilise-D Technical Validation Study (Dataset 3; n=18 subjects) and external validation with healthy young adults (Dataset 4; n=11 subjects) datasets. Outliers are excluded from plots for clarity but included in analyses. Overall model differences were assessed using a Friedman test ($p < 0.05$), followed by Wilcoxon signed-rank tests for pairwise comparisons. Horizontal lines indicate pairwise comparisons, with significance levels denoted above: one asterisk for $p < 0.05$, two asterisks for $p < 0.01$, three asterisks for $p < 0.001$, four asterisks for $p < 0.0001$; "ns" denotes not significant. Left column: internal validation; right column: external validation. **A, B** MAE in cm/s; **C, D** RMSE in cm/s; **E, F** MAPE in %; **G, H** R^2; **I, J** ICC(2,1).

Figure 3. Correlation between predicted and true gait speeds for ElderNet compared to benchmark models. Scatter plots showing predicted gait speed versus true gait speed, with each point representing a 10-second window. The black dashed line represents perfect agreement ($y = x$). Performance metrics are displayed in each panel: MAE, RMSE, MAPE, R^2, and ICC. Error and performance metrics are calculated at the window level. The top row shows results on the internal validation dataset (Dataset 3, n=18): **A)** ElderNet, **B)** State-of-the-art model, **C)** Supervised model. The bottom row shows results on the external dataset (Dataset 4, n=11): **D)** ElderNet, **E)** State-of-the-art model, **F)** Supervised model. ElderNet shows superior predictive performance across both cohorts.

Figure 4. Bland-Altman plot comparing gait speed predictions to ground truth in the Mobilise-D TVS test set (Dataset 3; n=18). Each dot represents a walking bout with a minimum duration of 10 seconds. The horizontal dashed lines indicate the 95% confidence intervals around the mean error. The red line represents the mean error of a lower-back-based highemodel evaluated on the entire Mobilise-D dataset (n=82) [24].

Figure 5. Receiver operating characteristic (ROC) curves for classifying mobility disability in the RUSH Memory and Aging project (Dataset 2; n = 423). Curves represent mean values (±1 standard deviation across 10 folds) obtained from nested cross-validation. Three different feature sets were compared: supervised-based accelerometry features (blue) extracted from standardized walking tests (8-ft and 32-ft walks) measuring gait speed, step length, cadence, and

ARTICLE IN PRESS

stride regularity; physical activity features (red) and gait quality features (green) extracted from daily-living accelerometer data. Classification performance, indicated by the area under the curve (AUC), improved from supervised-based features (AUC = 0.67 ± 0.09) to physical activity features (AUC = 0.69 ± 0.06), and was highest when using all available wrist multi-day gait metrics (AUC = 0.80 ± 0.06). A Friedman test showed a significant effect of feature set on AUC ($p < 0.001$). Post hoc Wilcoxon tests revealed the gait quality model outperformed both supervised ($p < 0.01$) and physical activity models ($p < 0.01$), with no significant difference between the latter two. Note that these results are based on the application of both stages of the ElderNet pipeline: gait detection, followed by the estimation of wrist-derived gait features.

ARTICLE IN PRESS

ARTICLE IN PRESS

Table 1. Characteristics of older adults in the Mobilise-D technical validation study (Dataset 3).

	Healthy Older Adults	**Congestive Heart Failure**	**Chronic Obstructive Pulmonary Disease**	**Parkinson's Disease**	**Proximal Femoral Fracture**	**Multiple Sclerosis**
No. of participants	17	11	17	15	11	14
Age (years)	72.4 ± 6.0	69.1 ± 11.7	69.3 ± 9.1	69.2 ± 7.5	79.6 ± 6.8	46.5 ± 8.7
Sex (M:F)	10:7	7:4	9:8	12:3	6:5	6:8
Montreal Cognitive Assessment score (0–30)	28.1 ± 1.3	26.9 ± 2.9	24.6 ± 3.3	23.9 ± 4.5	23.8 ± 4.3	26.2 ± 3.1
CAT score (0–40)	N/A	N/A	19.65 ± 8.95	N/A	N/A	N/A
FEV1 (L)	N/A	N/A	1.58 ± 0.58	N/A	N/A	N/A
Hoehn and Yahr stage (N)	N/A	N/A	N/A	I: 3, II: 7, III: 5	N/A	N/A
MDS-UPDRS III	N/A	N/A	N/A	30.67 ± 13.33	N/A	N/A
SPPB (0–12)	N/A	N/A	N/A	N/A	7.73 ± 3.10	N/A
Days since injury	N/A	N/A	N/A	N/A	132.1 ± 122	N/A
EDSS (0-6)	N/A	N/A	N/A	N/A	N/A	3.86 ± 1.60
No. of valid walking bouts per recording	40.41 ± 11.99	33.54 ± 22.83	29.35 ± 13.64	23.46 ± 17.29	26.82 ± 21.16	19.93± 13.37
Average length of walking bouts (s)	41.25± 31.68	36.75± 12.22	27.65± 11.56	45.18± 32.75	37.96± 17.60	36.77± 18.31

ARTICLE IN PRESS

ARTICLE IN PRESS

Average gait speed (cm/s)	72.38 ± 23.18	86.66 ± 22.54	64.42 ± 10.59	68.25± 26.88	59.61 ± 11.03	69.97 ± 15.91

CAT: chronic obstructive pulmonary disease assessment test, FEV1: forced expiratory volume in 1 s, MDS-UPDRS III: Movement Disorder Society Unified Parkinson Disease Rating Scale Part III, EDSS: Expanded Disability Status Scale., SPPB: short physical performance battery; scores on this test range from 0 to 12 (best performance) and values less than 10 indicates one or more mobility limitations and lower SPPB scores have been associated with an increased risk of disability[64]. The number of walking bouts, their average length, and gait speed were derived from the INDIP reference system during approximately 2.5 hours of continuous recording.

Table 2. Summary of datasets used in this study.

Dataset #	**Name**	**Sample Size**	**Population Characteristics**	**Duration**	**Purpose in Study**	**Annotation type**
Dataset 1	UK Biobank	>100,000	General population	7 days	SSL pretraining	Unlabeled
Dataset 2	RUSH MAP	819	Older adults	10 days	SSL adaptation; Clinical validation	Mobility disability scores
Dataset 3	Mobilise-D TVS	85	Older adults & gait impaired	2.5 hours	Fine-tuning & evaluation	Spatial-temporal gait annotations

ARTICLE IN PRESS

Dataset 4	Healthy young adults (Mobilise-D)	11	Healthy young adults	2.5 hours	External validation	Spatial-temporal gait annotations

Table 3. ElderNet performance for gait speed across different cohorts in the Mobilise-D TVS (Dataset 3) test set (n=18).

Cohort	MAE (cm/s)	RMSE (cm/s)	MAPE (%)	R^2	ICC
Healthy Adults	10.5	13.3	11	0.87	0.92
Congestive Heart Failure	9.7	14.0	13	0.79	0.90
Chronic Obstructive Pulmonary Disease	6.6	9.5	11	0.76	0.88
Parkinson's Disease	8.0	10.0	11	0.86	0.93
Proximal Femoral Fracture	7.5	9.4	13	0.67	0.87
Multiple Sclerosis	12.0	14.4	17	0.40	0.74

Performance is reported at the 10-second window level for each cohort.

Table 4. Comparison of ElderNet and a supervised model for estimating gait quality metrics in the Mobilise-D TVS (Dataset 3) test set.

Measure	Unit	Facet of Mobility	Metric	ElderNet (median [25th-75th])	Supervised (median [25th-75th])	p-value
Cadence	steps/minute	Rhythm	MAE	3.11 [2.26 3.88]	3.77 [2.94 4.46]	**<0.001*** **
			RMSE	5.19 [3.79 6.48]	5.83 [4.72 6.88]	**0.045***
			MAPE	3.71	4.26	**<0.01****

ARTICLE IN PRESS

				[2.54 4.54]	[3.27 5.39]	
			R^2	0.65 [0.58 0.74]	0.55 [0.45 0.74]	0.07
			ICC	0.83 [0.75 0.87]	0.77 [0.69 0.86]	**0.02***
Stride Length	cm	Pace	MAE	11.61 [8.83 13.68]	14.57 [10.90 17.43]	**<0.01****
			RMSE	14.09 [11.13 16.59]	18.30 [14.47 20.64]	**<0.001*****
			MAPE	11.09 [8.43 14.39]	15.62 [12.56 21.54]	**<0.001*****
			R^2	0.48 [0.14 0.69]	0.31 [-0.62 0.45]	**<0.001*****
			ICC	0.72 [0.59 0.83]	0.54 [0.35 0.64]	**<0.001*****
Stride Regularity	A.U. (0-1)	Variability	MAE	0.08 [0.08 0.10]	0.08 [0.07 0.09]	0.30
			RMSE	0.11 [0.10 0.13]	0.11 [0.09 0.12]	0.21
			MAPE	28.25 [23.25 36.36]	31.27 [19.45 35.82]	0.30
			R^2	0.55 [0.38 0.69]	0.67 [0.42 0.76]	0.30
			ICC	0.76 [0.68 0.82]	0.84 [0.75 0.88]	**0.012**

Performance metrics are presented as median [25th–75th percentile]. The test set included 18 participants from the Mobilise-D dataset. Statistical significance between models (p-values) was determined using the Wilcoxon signed-rank test for dependent samples. Bolded values with an asterisk indicate statistically significant differences after Bonferroni correction for multiple comparisons. one asterisk for $p < 0.05$, two asterisks for $p < 0.01$, three asterisks for $p < 0.001$, four asterisks for $p < 0.0001$; A.U., arbitrary unit. See Table S1 for the parallel results for gait speed.

Table 5. Characteristics of the RUSH Memory and Aging project cohort (Dataset 2)

Variable	All Participants (n=787*)	Subset with Supervised Assessment (n=423)
Demographics		
Age, years (mean ± SD)	83.3 ± 7.4	83.6 ± 6.9
Gender (M:F)	606:181	324:99
Mobility disability status (Yes:No)	358:429	176:247
Daily living gait‡		

ARTICLE IN PRESS

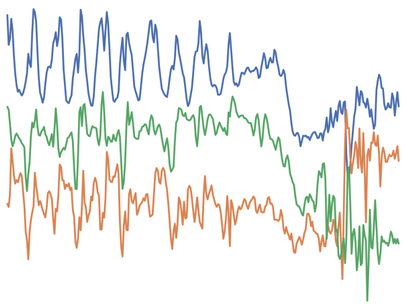
Wear days (median [IQR])	9 [9 10]	9 [9 10]
Walking bouts per day (median [IQR])	76.9 [41.53 106.9]	80.6 [51.0 108.2]
Median bout duration (sec; median [IQR])	16 [15 17]	16 [15 17]
Median daily living gait speed (cm/sec; median [IQR])	65.9 [57.0 74.2]	68.2 [59.6 75.9]
Median daily living stride length (cm; median [IQR])	97.7 [90.8 104.2]	98.5 [92.3 104.6]
Median daily living cadence (steps/minute; median [IQR])	97.3 [90.8 105.1]	98.1 [92.3 105.4]
Median daily living stride regularity (a.u; median [IQR])	0.31 [0.27 0.36]	0.31 [0.28 0.35]
Supervised Assessments		
Average gait speed 32-foot test (cm/sec; median [IQR])	N/A	47.2 [42.2 52.1]
Average gait speed 8-foot test (cm/sec; median [IQR])	N/A	49.8 [43.2 52.64]
Average stride length 32-foot test (cm; median [IQR])	N/A	97.5 [81.3 114.74]
Average stride length 8-foot test (cm; median [IQR])	N/A	101.6 [81.31 122.0]
Average cadence 32-foot test (steps/minute; median [IQR])	N/A	92.3 [82.9 101.5]
Average cadence 8-foot test (steps/minute; median [IQR])	N/A	93.9 [85.1 101.2]
Average stride regularity 32-foot test (a.u.; median [IQR])	N/A	0.61 [0.47 0.70]
Average stride regularity 8-foot test (a.u.; median [IQR])	N/A	0.56 [0.40 0.69]

*From the initial cohort of 819 participants, only 787 had sufficient data (defined as more than 3 valid days of accelerometer wear) and were included in this analysis.

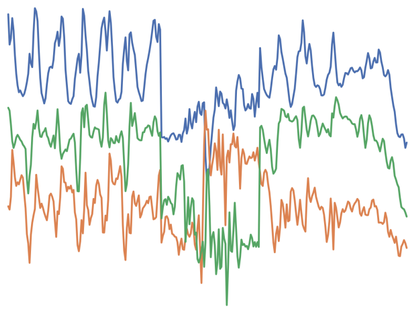
‡Daily-living gait metrics were calculated for each walking bout. For each participant, the median value across bouts was determined, and the cohort-level values are presented as the median [IQR] of those participant-level medians. IQR, interquartile range; A.U., arbitrary unit.

Large, daily living, unlabeled dataset

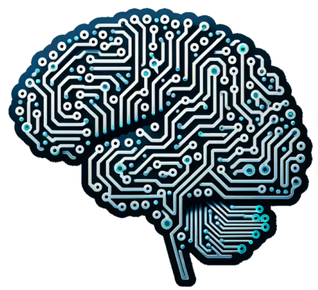
Acceleration Signal



Augmented Signal



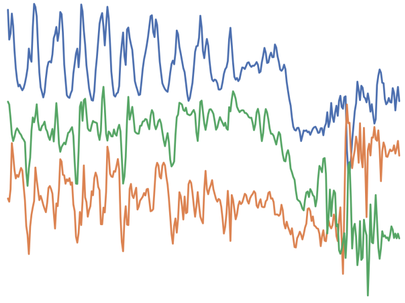
Foundation Model



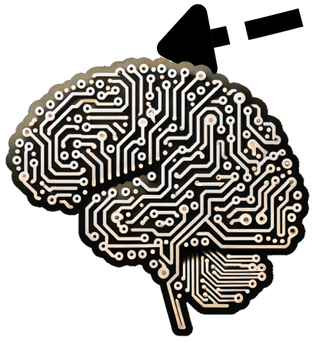
Feature Vector

Smaller, daily living, labeled dataset

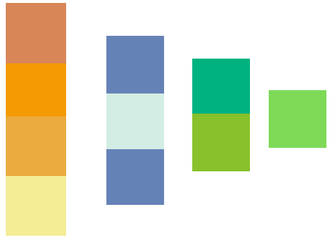
Acceleration Signal



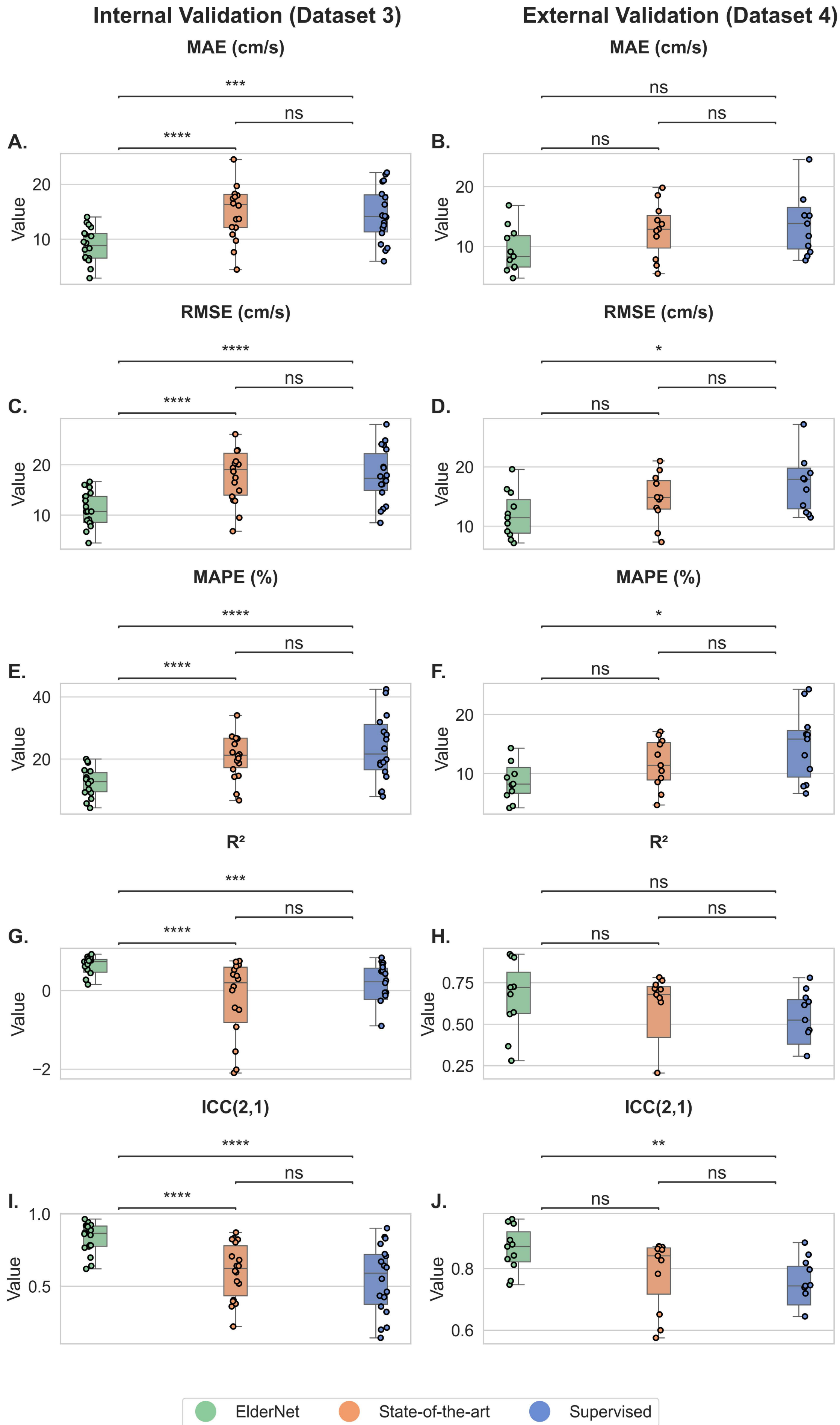
Fine-Tuning



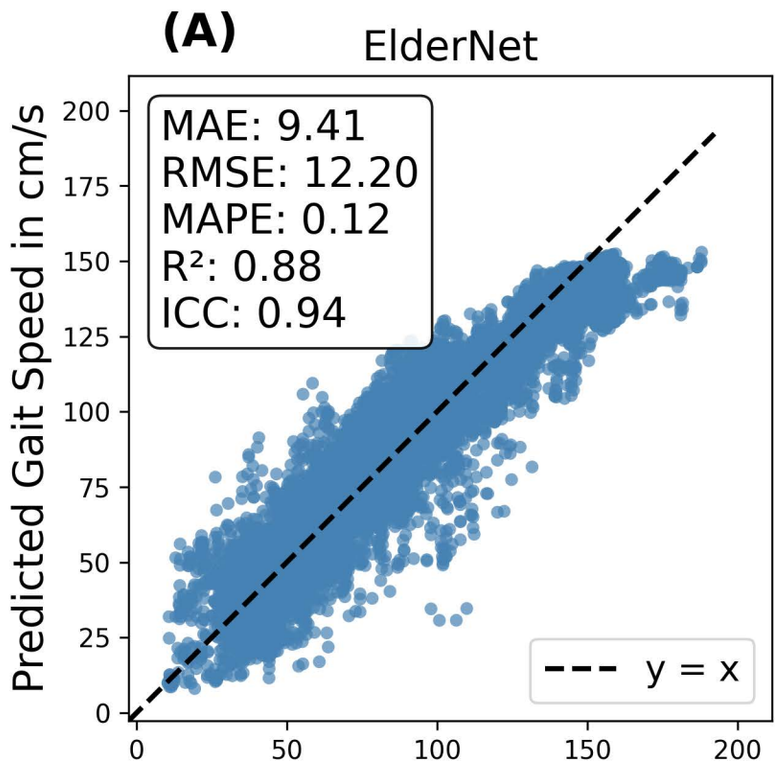
Regression Head

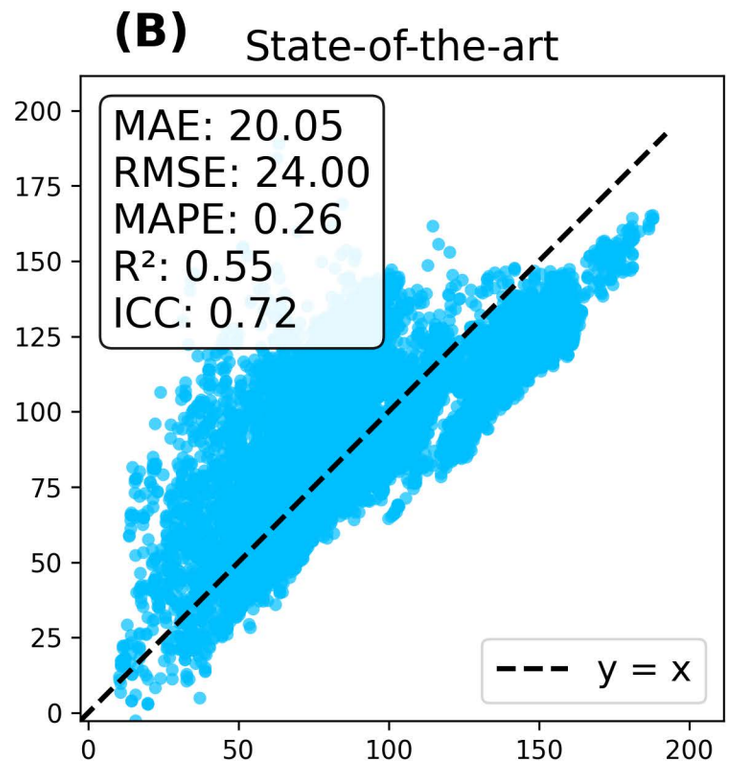


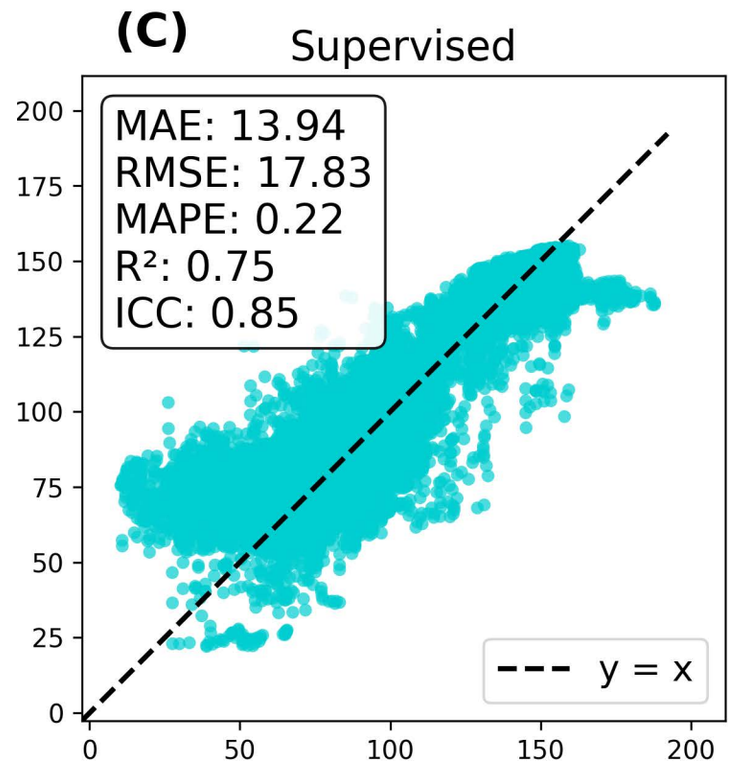
Gait Speed Estimation


Internal Validation (Dataset 3)
External Validation (Dataset 4)
A.
B.
MAE (cm/s)
***
ns
****
ns
ns
ns
Value
20
10
RMSE (cm/s)
C.
D.
****
ns
****
*
ns
ns
MAPE (%)
E.
F.
****
ns
****
40
20
*
ns
ns
10
R²
G.
H.
***
ns
****
ns
ns
0
−2
0.75
0.50
0.25
ICC(2,1)
I.
J.
****
ns
****
**
ns
ns
1.0
0.5
0.8
0.6
ElderNet
State-of-the-art
Supervised

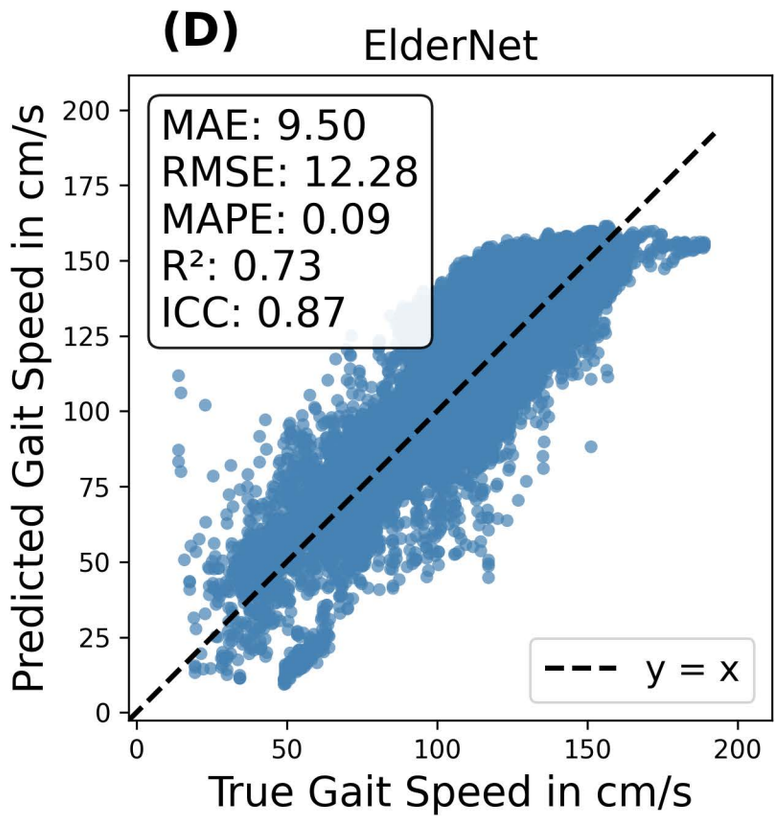
Internal Validation (Dataset 3)

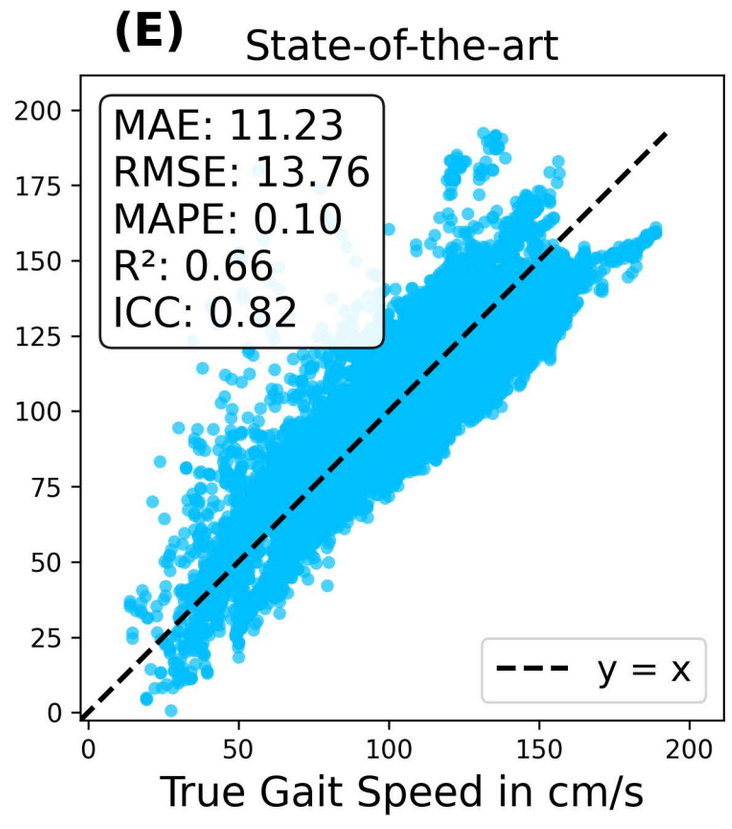


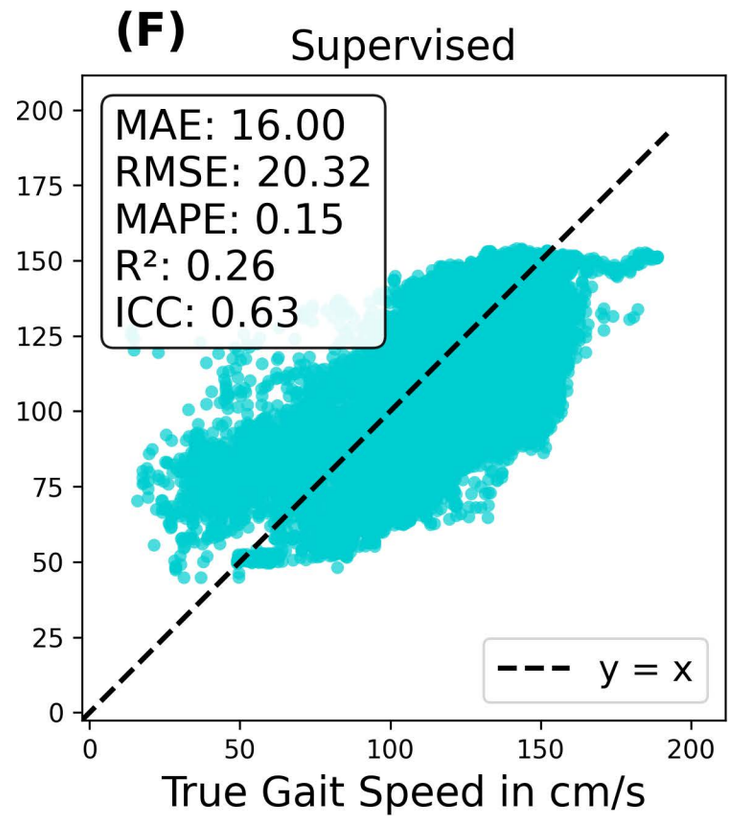


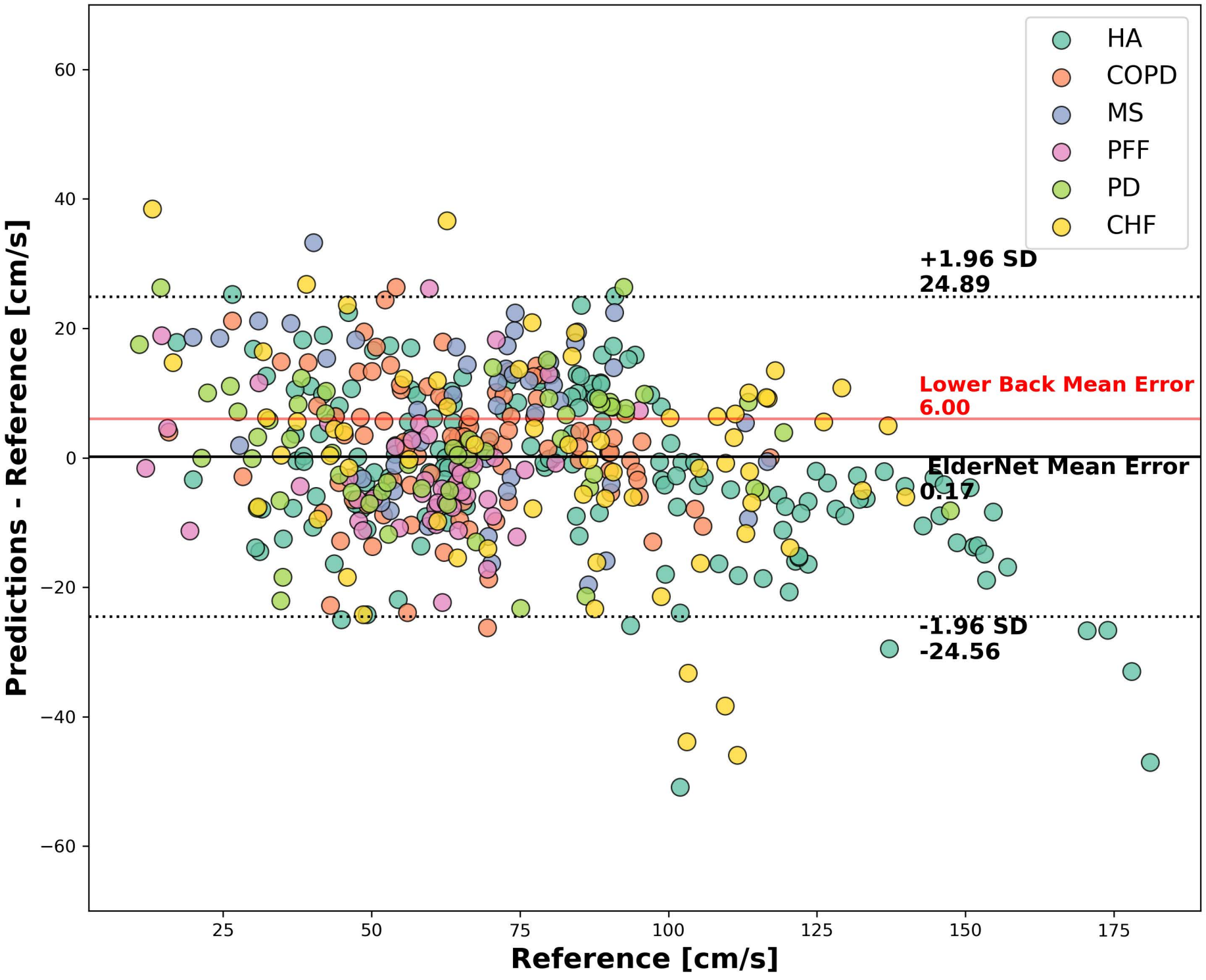


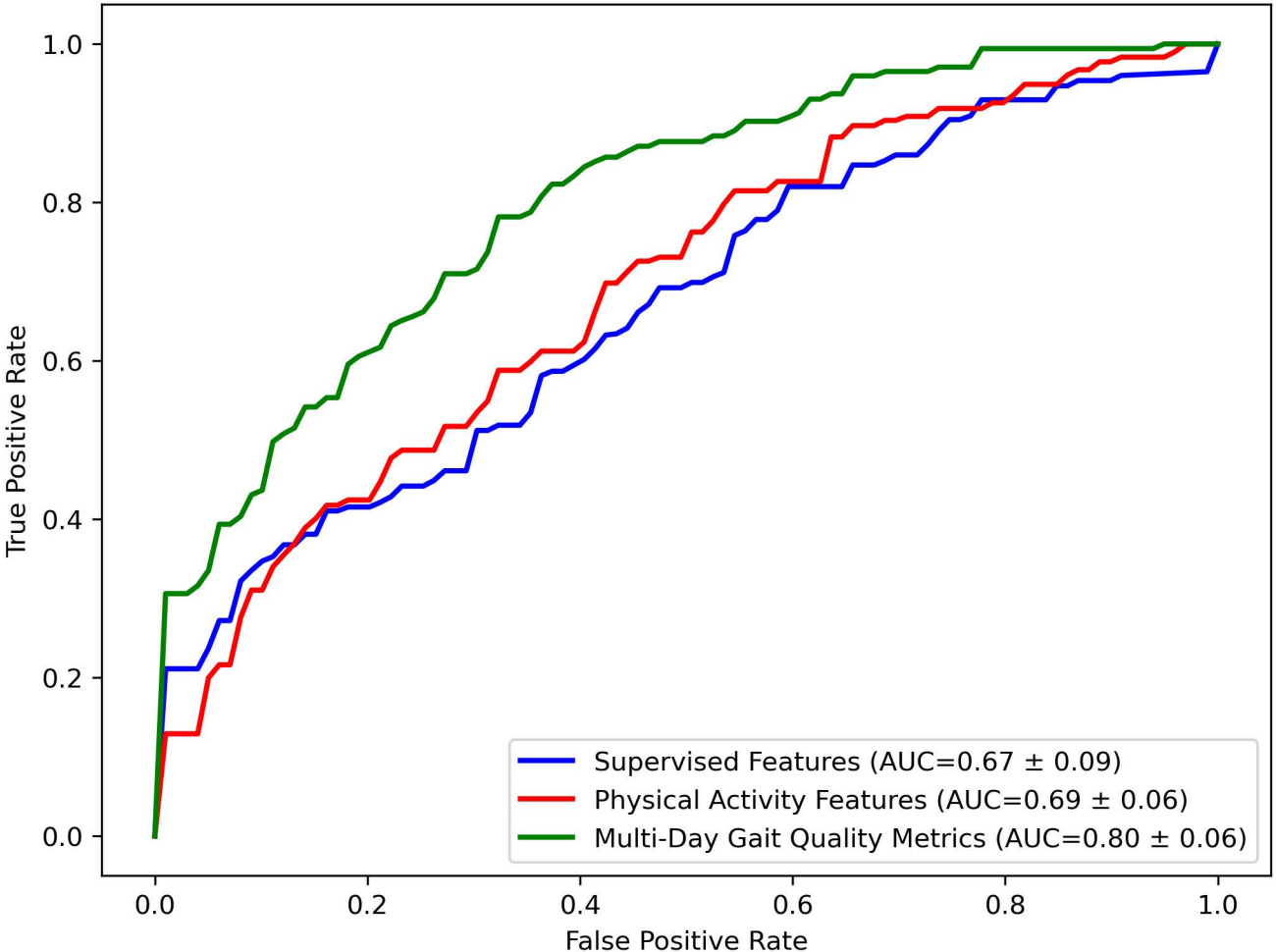
External Validation (Dataset 4)








HA
COPD
MS
PFF
PD
CHF
+1.96 SD
24.89
Lower Back Mean Error
6.00
ElderNet Mean Error
0.17
-1.96 SD
-24.56
Predictions - Reference [cm/s]
Reference [cm/s]
60
40
20
0
−20
−40
−60
25
50
75
100
125
150
175


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Supervised Features (AUC=0.67 ± 0.09)
Physical Activity Features (AUC=0.69 ± 0.06)
Multi-Day Gait Quality Metrics (AUC=0.80 ± 0.06)