Identification of Fatigue and Sleepiness in Immune and Neurodegenerative Disorders from Measures of Real-World Gait Variability

Chloe Hinchliffe[1], Rana Zia Ur Rehman[2], Diogo Branco[3], Dan Jackson[4], Teemu Ahmaniemi[5], Tiago Guerreiro[3], Meenakshi Chatterjee[6], Nikolay V. Manyakov[7], Ioannis Pandis[6], Kristen Davies[1], Victoria Macrae[8], Svenja Aufenberg[9], Emma Paulides[10], Hanna Hildesheim[11], Jennifer Kudelka[11], Kirsten Emmert[11], Geert Van Gassen[12], Lynn Rochester[1], C. Janneke van der Woude[10], Ralf Reilmann[9], Walter Maetzler[11], Wan-Fai Ng[8], Silvia Del Din[1], IDEA-FAST Consortium

***Abstract*— Current assessments of fatigue and sleepiness rely on patient reported outcomes (PROs), which are subjective and prone to recall bias. The current study investigated the use of gait variability in the "real world" to identify patient fatigue and daytime sleepiness. Inertial measurement units were worn on the lower backs of 159 participants (117 with six different immune and neurodegenerative disorders and 42 healthy controls) for up to 20 days, whom completed regular PROs. To address walking bouts that were short and sparse, four feature groups were considered: sequence-independent variability (SIV), sequence-dependant variability (SDV), padded SDV (PSDV), and typical gait variability (TGV) measures. These gait variability measures were extracted from step, stride, stance, and swing time, step length, and step velocity. These different approaches were compared using correlations and four machine learning classifiers to separate low/high fatigue and sleepiness.**

Most balanced accuracies were above 50%, the highest was 57.04% from TGV measures. The strongest correlation was 0.262 from an SDV feature against sleepiness. Overall, TGV measures had lower correlations and classification accuracies.

Identifying fatigue or sleepiness from gait variability is extremely complex and requires more investigation with a larger data set, but these measures have shown performances that could contribute to a larger feature set.

***Clinical relevance*— Gait variability has been repeatedly used to assess fatigue in the lab. The current study, however, explores gait variability for fatigue and daytime sleepiness in real-world scenarios with multiple gait-impacted disorders.**

[1]Chloe Hinchliffe (corresponding author), Kristen Davies, Lynn Rochester, and Silvia Del Din are with the Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom. chloe.hinchliffe@newcastle.ac.uk, kristen.davies@newcastle.ac.uk, lynn.rochester@newcastle.ac.uk, and silvia.del-din@newcastle.ac.uk

[2]Rana Zia Ur Rehman is with Janssen Research & Development, High Wycombe, United Kingdom. rrehman5@its.jnj.com

[3]Diogo Branco and Tiago Guerreiro are with LASIGE, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal. djbranco@fc.ul.pt and tjvg@di.fc.ul.pt

[4]Dan Jackson is with Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom. dan.jackson@newcastle.ac.uk

[5]Teemu Ahmaniemi is with VTT, Visiokatu 4, 33720 Tampere, Finland. Teemu.Ahmaniemi@vtt.fi

[6]Meenakshi Chatterjee and Ioannis Pandis are with Janssen, Cambridge, United States of America. mchatte4@ITS.JNJ.com and ipandis@its.jnj.com

[7]Nikolay V. Manyakov is with Janssen Research & Development, Beerse, Belgium nmanyak1@its.jnj.com

[8]Victoria Macrae and Wan-Fai Ng are with NIHR Newcastle Clinical Research Facility, Newcastle, United Kingdom. victoria.macrae@newcastle.ac.uk and wan-fai.ng@newcastle.ac.uk

[9]Svenja Aufenberg and Ralf Reilmann are with the George-Huntington-Institute, Muenster, Germany. svenja.aufenberg@ghi-muenster.de and ralf.reilmann@ghi-muenster.de

[10]Emma Paulides and C. Janneke van der Woude are with the Department of Gastroenterology and Hepatology, Erasmus University Medical Center, Molewaterplein 40, 3015 GD Rotterdam, The Netherlands. e.paulides@erasmusmc.nl and c.vanderwoude@erasmusmc.nl

[11]Hanna Hildesheim, Jennifer Kudelka, Kirsten Emmert, and Walter Maetzler are with the Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany. hannaluise.hildesheim@uksh.de, Jennifer.Kudelka@uksh.de, k.emmert@neurologie.uni-kiel.de, and w.maetzler@neurologie.uni-kiel.de

[12]Geert Van Gassen is with Medical Affairs, Takeda, Brussels, Belgium. geert.vangassen@takeda.com

I. INTRODUCTION

Fatigue and sleep disruptions are symptoms commonly reported in individuals with neurological and immune disorders. For instance, over 80% of those with systemic lupus erythematosus (SLE) suffer from abnormal fatigue [1] and fatigue in Parkinson's disease (PD) has a reported prevalence rate ranging from 33% to 58% [2].

A common method for assessing daytime sleepiness and fatigue is collecting patient reported outcomes (PRO)s, where the patients complete questionnaires and diaries designed to record how the patient is feeling. PROs, however, are subjective, prone to recall bias, and do not capture short-term changes over time. Furthermore, evidence shows that individuals who are sleep deprived are prone to underestimating their fatigue-related impairments [3].

Wearable sensors would negate these downsides since they would give objective, continuous reports on the patient's physiological state. Inertial measurement units (IMU)s (triaxial accelerometer and gyroscope) are becoming an increasingly popular option for wearable technology, due the affordability and suitability for use in patients' homes.

Gait variability assesses step-to-step fluctuations and, in data collected in the lab from healthy older adults, has been shown to be impacted by physical fatigue [4], mental

fatigue [5], and sleepiness [6]. However, for personal or long term use, gait variability must be usable in "free living" environments. This introduces complications since natural day-to-day bouts of walking are typically short and sparse, especially for those with limited mobility. As such, the current literature investigating gait variability mostly takes place in a laboratory or hospital and the participants are given a specific task to perform. A 2018 literature review for gait analysis outside the lab found only 18 studies that analysed the standard deviation (SD) and/or coefficient of variation (CoV) of gait characteristics, predominantly step time [7]. A 2022 study found that gait is largely variable in the real world and also noted a lack of literature examining natural gait variability [8]. Furthermore, to our knowledge, no studies have specifically investigated the impact of fatigue or sleepiness in adults on walking gait variability outside of laboratory- or task-based procedures.

The current study aims to investigate the use of real-world gait variability for identifying physical fatigue, mental fatigue, and daytime sleepiness. Four feature groups individually designed to overcome the short and sparse nature of walking found in natural behaviour were compared using correlations and popular machine learning classifiers.

II. MATERIALS AND METHODS

The data used in this analysis were collected at four sites in Europe as part of the IDEA-FAST feasibility study [9]. All participants provided informed consent and the study was approved by all local ethics committees. The participants included healthy controls (HC=42) and six different neurodegenerative disorders (NDD) and immune-mediated inflammatory diseases (IMID): Parkinson's disease (PD=25), Huntington's disease (HD=14), rheumatoid arthritis (RA=24), systemic lupus erythematosus (SLE=18), primary Sjogren's syndrome (PSS=18), and inflammatory bowel disease (IBD=18). The 159 participants wore various sensors, 132 of which wore an IMU sensor on the lower back for four periods of five consecutive days at home. The IMU was a MoveMonitor device [10] with a sampling rate of 100 Hz and a range of $\pm$16g. During the trial period, the participants answered the PRO questionnaires up to four times daily. The physical fatigue (PF) and mental fatigue (MF) were self reported on a scale of 0-6 and daytime sleepiness (Karolinska's sleepiness scale (KSS) [11]) on 0-9.

Bouts of walking that began in the two hour period before each PRO submission were identified. Using the triaxial accelerometer data, six gait characteristics were extracted with validated methods [12]: step time, stride time, stance time, swing time, step length, and step velocity. The gait variability measures considered for extraction from these six gait micro characteristics were inspired by heart rate variability (HRV) analysis [13], [14]. Possible features were separated into four gait variability feature groups: sequence-independent variability (SIV), sequence-dependent variability (SDV), padded SDV (PSDV), and typical gait variability (TGV) measures. These four feature groups were designed to handle the problem of short, sparse walking bouts in different

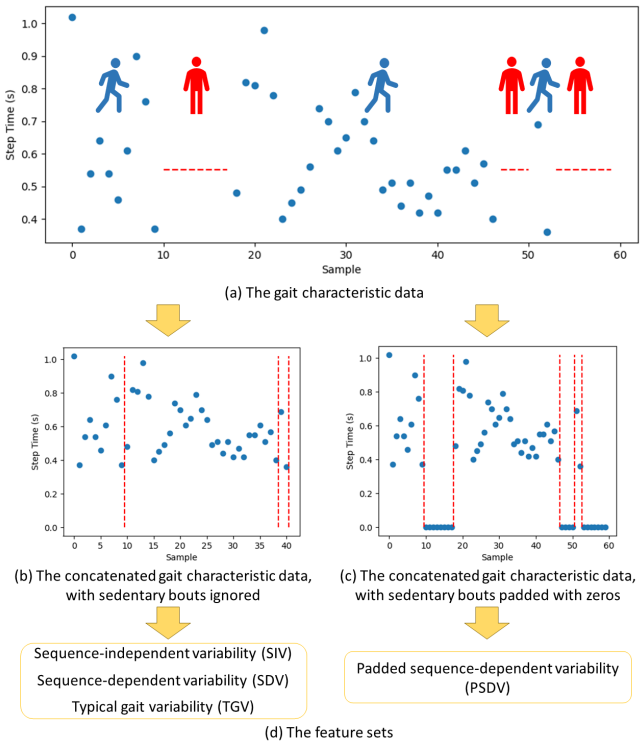

Fig. 1 *A flowchart that shows the protocol for each feature set to be compared. (a) gives an example of the gait characteristic data, where the blue dots represent the gait data (here, the example is step time) and the red, dashed lines represent when the participant is not walking. (b) gives the concatenated data, where non-walking periods are ignored. (c) gives the padded data, where the non-walking periods are replaced with zeros. (d) gives the feature sets extracted from each representation of the data.*

ways. SIV, SDV, and TGV measures simply treated the data as one continuous walking bout, ignoring the gaps where the participant was sedentary. SIV used only measures of variability where measures were independent of the data's sequence (order) and missing data did not introduce excess noise: 16 statistical (SD, CoV, 20th and 80th percentiles, successive differences analysis, etc.) and 23 Poincaré plot analysis (SD perpendicular to the line, cardiac sympathetic index, etc.) measures. Where successive differences were required, only the differences between steps within the same walking bout were extracted. SDV used measures that depended on the temporal sequence of the data: six complexity (three fractal dimensions (FD), Lempel-Ziv complexity (LZC), etc.), three entropy (spectral, sample, and multiscale sample (MSE)), and eight frequency analysis (frequency with maximum amplitude, spectral power of low/high frequencies, etc.) measures. Parameters such as frequency cut-offs and entropy inputs ($m = 2$, $r = 0.2 \times$SD) were defined by HRV [13]. PSDV extracted the same features as SDV, but padded sedentary bouts with zeros at a frequency of typical human walking (1/0.52 for step and stance, 1/1.04 for stride and swing characteristics [15]), thus additionally encoding the bout information (macro characteristics). TGV used the two variability measures commonly used in gait research (SD and

CoV) as a "control" approach for comparison. This protocol is outlined in Figure 1. Gait measures associated with a step time ≤ 0.25s or ≥ 1.25s were removed and if less than 30 seconds of walking data were associated with a PRO, that instance was not included. This resulted in 1771, 1729, and 1645 samples for PF, MF, and KSS, respectively, from 102 subjects. The number of input features for each of the six gait micros were TGV=12, SIV=228, SDV=96, and PSDV=96.

Once these features were extracted for each gait characteristic, they were classified as low fatigue/sleepiness (scores of 0-2) or high fatigue/sleepiness (scores of 3 or above) [16] with four popular machine learning classifiers: support vector machine (SVM) with a radial basis function kernel, k-nearest neighbours (kNN), random forest (RF), and Gaussian naive Bayesian network (BayesNet). The features were prepared by replacing missing values with the training-set median, then equalising the classes with synthetic minority oversampling technique (SMOTE) [17], and reducing the input dimension with principal component analysis (PCA). The data were split into train/test sets with 5-fold cross-validation, where 20% of the participants were excluded as the test data for each fold. The classification performance of the four feature groups were evaluated using balanced accuracy to prevent skewed results from the imbalanced classes, with the mean balanced accuracy of each fold reported in the results. The classifier-input preparation and machine learning was implemented with the `scikit-learn` package [18].

To investigate these features individually, their correlations with each PRO were explored. Repeated measures correlation (RMCorr) was used to avoid bias from the use of multiple samples from each subject. RMCorr removes the variance between participants, thus providing the best linear fit for each participant using parallel regression lines with varying y-intercepts [19]. The RMCorr coefficient is bounded by -1 to 1, with 0 indicating no association, and was implemented in python using `rm_corr` [20].

III. RESULTS

The highest balanced accuracy from each PRO was PF = 57.04% (TGV with step length and SVM), MF = 56.19% (PSDV with step velocity and SVM), and KSS = 56.31% (SIV with step time and SVM), and the majority of accuracies were above chance (50%). Tables 1 and 2 give summaries of the balanced accuracies returned from the classifiers: means of the gait micros and PROs, respectively.

Table 1 shows that step time, length, and velocity returned higher accuracies than stride and swing time, and stance time was the lowest performing characteristic with the lowest classifier robustness (higher SDs). Combining these measures together in 'all' did not improve or degrade the accuracies. TGV and PSDV were, overall, the lower performing groups but SIV had the lowest classifier robustness. Furthermore, SDV returned the highest accuracy for most gait characteristics. However, these outcomes should be considered critically, since the SDs are often considerably high compared to differences between the means. Overall, the highest mean accuracy was 54.30% from the SDV group with step velocity.

TABLE I *Mean balanced accuracies±SD of each gait characteristic for each feature group, averaged across the four classifiers and three PROs. Bold values denote the highest mean accuracy for each gait characteristic.*

Gait micro	TGV (%)	SIV (%)	SDV (%)	PSDV (%)
Step Time	52.34±0.4	53.03±1.0	**53.61**±0.3	52.48±0.6
Step Velocity	52.59±0.1	53.31±0.5	**54.30**±0.6	53.54±0.2
Step Length	52.04±1.2	52.93±0.8	**53.27**±0.3	52.03±0.5
Stance Time	48.78±0.8	48.72±1.1	50.14±0.8	**51.02**±1.1
Stride Time	50.82±0.3	51.35±0.9	**52.03**±0.6	51.64±0.4
Swing Time	50.96±1.0	52.05±0.6	**52.19**±0.4	51.72±0.4
All	51.57±0.3	52.72±0.7	**53.13**±0.5	51.72±0.4

TABLE II *Mean balanced accuracies±SD of each PRO for each feature group, averaged across the four classifiers and six gait characteristics. Bold values denote the highest mean accuracy for each PRO.*

PRO	TGV (%)	SIV (%)	SDV (%)	PSDV (%)
PF	51.73±0.9	51.10±0.5	**52.59**±0.6	51.62±0.5
MF	51.78±0.4	51.79±0.9	52.10±0.4	**52.57**±0.5
KSS	50.39±0.9	53.15±0.8	**53.31**±0.5	51.87±0.9

Table 2 shows that for PF and MF, when SD is taken into account, the differences between the means of the feature groups are negligible. The KSS, however, returned higher accuracies with the SIV and SDV groups than with TGV and PSDV (minimum difference of 1.28%). This indicates that the KSS PRO is less robust to classifier inputs.

The strongest RMCorr was 0.262 (LZC of step velocity with KSS PRO) and therefore, based on r^2, explains 6.9% of variability of KSS. Generally, the KSS returned the strongest positive and negative correlations, with less variation in the PF and MF PROs. When considering the correlations of the individual features (absolute RMCorr averaged across the PROs), for TGV the SD feature outperformed CoV for each gait micro and step velocity returned the strongest correlations for both measures. For SIV, the statistical features were generally outperformed by the Poincaré analysis, aside

TABLE III *Summary of RMCorrs for each feature group, including the percentage of features with $p<0.01$ with at least one PRO ('Sig. Features'), the mean of the positive (+r) and negative (-r) correlations, and the minimum and maximum of the correlations (r). Bold denotes the 'best' group for each correlation metric.*

Outcome	TGV	SIV	SDV	PSDV
Sig. Features	50%	53%	58%	**70%**
Mean +r	0.038	0.048	**0.072**	0.060
Max +r	0.122	0.162	**0.262**	0.177
Mean -r	-0.010	-0.037	-0.050	**-0.069**
Min -r	-0.016	-0.174	-0.185	**-0.243**

from the 80th percentile measure which returned three of the group's five best features. For the SDV, the MSE, LZC, and spectral entropy returned most of the strongest correlations. For PSDV, the MSE, Katz FD, and correlation FD (CD) were the best. With padding, all CD measures' correlations improved but, for all LZC they decreased. Furthermore, the frequency analysis was generally outperformed by measures of complexity. Table 3 reports a summary of the RMCorr outcomes. The table shows that TGV was outperformed by the other three feature groups, and SDV and PSDV returned the strongest correlations. Furthermore, padding the gait micros improved the number of statistically significant features ($p < 0.01$) to 70% associated with at least one PRO.

IV. DISCUSSION

The majority of the correlations and classification accuracies in this analysis were low (<0.3 correlation, <60% accuracy), which indicates to the complexity of this task. It is possible that in real world scenarios, when people are tired or fatigued it will not significantly impact how they walk. Naturally, identifying mental states from physical manifestations is challenging, but even low associations can have clinical relevance. However, the main focus of this particular study was to compare different approaches to handling the complications of real-world gait variability.

Overall, the SDV feature group seemed to be the best, with higher and robust classification accuracies and stronger correlations. Padding the SDV improved the percentage of statistically significant features, but failed to improve the overall correlations and classifier performances. Generally, TGV was slightly outperformed in classification performances and it returned distinctly lower correlations. However, it returned the highest balanced accuracy in this analysis. Therefore, SD and CoV may be useful for binary classification, but to capture more nuance in PRO associations, more advanced gait variability measures are required.

The main limitation of this analysis is that this is a feasibility study with a fairly small cohort and very sparse data. It is also unknown what the impacts of the different disorders are, for instance HD can impact gait variability [21], since these individual cohorts are too small for reliable analysis. Furthermore, using parameters such as frequency cutoffs based on HRV may have restricted capture of gait specific information, therefore requiring further investigation.

In conclusion, objectively identifying fatigue in the real world is a challenging task and exploring advanced measures of gait variability can provide more insight to fatigue and daytime sleepiness than SD and CoV alone.

ACKNOWLEDGEMENT

This research was funded by the IDEA-FAST project, which has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No. 853981. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and associated partners.

This work was also supported by the National Institute for Health Research (NIHR) Newcastle Biomedical Research Centre (BRC) based at Newcastle Upon Tyne Hospital NHS Foundation Trust and Newcastle University; and the NIHR/Wellcome Trust Clinical Research Facility (CRF) infrastructure at Newcastle upon Tyne Hospitals NHS Foundation Trust. All opinions are those of the authors and not the funders. Neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein.

REFERENCES

[1] C. M. Tench, I. McCurdie, P. D. White, and D. P. D'Cruz, "The prevalence and associations of fatigue in systemic lupus erythematosus," *Rheumatology (Oxford)*, vol. 39, no. 11, pp. 1249–1254, 2000.

[2] J. H. Friedman *et al.*, "Fatigue in Parkinson's disease: A review," *Movement Disorders*, vol. 22, no. 3, pp. 297–308, Feb. 2007.

[3] X. Zhou *et al.*, "Mismatch between subjective alertness and objective performance under sleep restriction is greatest during the biological night," *J Sleep Res*, vol. 21, no. 1, pp. 40–49, Feb. 2012.

[4] G. Zhang *et al.*, "Identifying fatigue indicators using gait variability measures: A longitudinal study on elderly brisk walking," *Sensors (Switzerland)*, vol. 20, no. 23. MDPI AG, pp. 1–12, Dec. 01, 2020.

[5] M. Behrens *et al.*, "Mental Fatigue Increases Gait Variability During Dual-task Walking in Old Adults," *The Journals of Gerontology: Series A*, vol. 73, no. 6, pp. 792–797, May 2018.

[6] S. Nakakubo *et al.*, "The Association Between Excessive Daytime Sleepiness and Gait Parameters in Community-Dwelling Older Adults: Cross-Sectional Findings From the Obu Study of Health Promotion for the Elderly," *J Aging Health*, vol. 30, no. 2, pp. 213–228, Feb. 2018.

[7] L. C. Benson, C. A. Clermont, E. Bošnjak, and R. Ferber, "The use of wearable devices for walking and running gait analysis outside of the lab: A systematic review," *Gait Posture*, vol. 63, pp. 124–138, 2018.

[8] L. Baroudi, X. Yan, M. W. Newman, K. Barton, S. M. Cain, and K. A. Shorter, "Investigating walking speed variability of young adults in the real world," *Gait Posture*, vol. 98, pp. 69–77, Oct. 2022.

[9] "IDEA-FAST." [Online]. Available: https://idea-fast.eu/.

[10] McRoberts, "McRoberts - Moving Technology — To monitor physical activity and performance." https://www.mcroberts.nl/.

[11] T. Åkerstedt and M. Gillberg, "Subjective and Objective Sleepiness in the Active Individual," *Int. J. Neurosci.*, vol. 52, pp. 29–37, 2009.

[12] S. Del Din, A. Godfrey, and L. Rochester, "Validation of an Accelerometer to Quantify a Comprehensive Battery of Gait Characteristics in Healthy Older Adults and Parkinson's Disease: Toward Clinical and at Home Use," *IEEE J Biomed Health Inform*, vol. 20, no. 3, pp. 838–847, May 2016.

[13] D. Makowski *et al.*, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behav Res Methods*, vol. 53, no. 4, pp. 1689–1696, Feb. 2021.

[14] T. Pham, Z. J. Lau, S. H. A. Chen, and D. Makowski, "Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial," *Sensors*, vol. 21, no. 12, p. 3998, Jun. 2021.

[15] S. Voss *et al*., "Normative database of spatiotemporal gait parameters using inertial sensors in typically developing children and young adults," *Gait Posture*, vol. 80, pp. 206–213, Jul. 2020.

[16] S. J. Bowman *et al.*, "Measurement of fatigue and discomfort in primary Sjögren's syndrome using a new questionnaire tool," *Rheumatology*, vol. 43, no. 6, pp. 758–764, Jun. 2004.

[17] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning," *J. Mach. Learn. Res.*, vol. 18, no. 17, pp. 1–5, 2017.

[18] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[19] J. Z. Bakdash and L. R. Marusich, "Repeated measures correlation," *Front. Psychol.*, vol. 8, no. MAR, p. 456, Apr. 2017.

[20] R. Vallat, "Pingouin: statistics in Python," J. *Open Source Softw.*, vol. 3, no. 31, p. 1026, Nov. 2018.

[21] Y. Moon, J. H. Sung, R. An, M. E. Hernandez, and J. J. Sosnoff, "Gait variability in people with neurological disorders: A systematic review and meta-analysis," *Hum. Mov. Sci.*, vol. 47, pp. 197–208, Jun. 2016.