Digital mobility outcomes as an early marker of disease progression in multiple sclerosis

Lucy Cheesman[1], John Mitchell[1], Gowtham Konduri[1], Gavin Brittain[2], Ellen Buckley[1], Lynn Rochester[3], Brian Caulfield[4], Silvia Del Din[3], Clemens Becker[5], Carl-Philipp Jansen[5], Walter Maetzler[6], Alison Yarnall[3], Judith Garcia-Aymerich[7], Alícia Josa-Culleré[7], Gloria Dalla Costa[8], Gaßner Heiko[9], Klarissa Stuerner[6], Beatrix Vereijken[10], Clint Hansen[11], Letizia Leocani[12], Basil Sharrack[2], Vita Lanfranchi[1], Shaoxiong Sun[1]

I. Introduction

Multiple sclerosis (MS) is the most common cause of neurological disability in young adults and is increasing in incidence [1]. Clinical trials and disease-modifying therapies in MS commonly target disability progression. One key metric is "no evidence of progression" (NEP), a composite measure derived from in-clinic scores for three functional tests. However, there are a number of criticisms of these methods, including non-linearity and poor sensitivity to change [2]. Therefore, there is an incentive to develop more granular methods to monitor disease progression, with particular focus on early detection [1, 3].

II. Objective

This work aims to explore the utility of aggregate digital mobility outcomes (DMOs) derived from wearable inertial measurement unit (IMU) devices as a marker of disease progression in people with multiple sclerosis (pwMS).

III. Methods

The Mobilise-D clinical validation study dataset consists of DMOs and associated clinical measures, including NEP status, for 602 pwMS. Data was collected at five timepoints (t1-5) across a two-year period at 6-monthly intervals, using either Axivity AX6 or McRoberts MM+ IMUs, worn on the lower back for a 7-day "free-living" collection period. Data was processed to extract 24 weekly aggregate DMOs at each timepoint [4].

From a total potential set of 3,010 observations, 843 were missing one or more DMOs. Observations with missing data were removed, leaving 2167 valid samples across 580 participants. Absolute and relative change compared to the previous timestamp was calculated for all DMOs at t2-t5, giving a total of 72 input features. Data was split participant-wise into train and test sets.

Principal component analysis (PCA) was applied to reduce dimensionality of highly correlated features. In total, 25 principal components were extracted, representing 95% of variance. Four target variables were identified:

1) NEP status at current timestamp (t)
2) NEP status at t+1
3) Change in NEP status from t-1 to t
4) Change in NEP status from t to t+1.

ML algorithms (random forest, logistic regression (LR), gradient boosting, and support vector machine (SVM)) were implemented, using Bayesian optimisation to tune hyperparameters. Models were evaluated using macro averaged recall in order to emphasise performance in the minority class (failure of NEP criteria) as well as accuracy and F1 score. Feature importance analysis was undertaken on the trained models.

IV. Results & Discussion

Models trained on the top 25 principal components outperformed those trained on the original features and deltas. Top performing models are as follows:

1) **NEP status at t**
LR: accuracy 0.68, macro-recall 0.70, F1 0.58
2) **NEP status at t+1**
SVM: accuracy 0.71, macro-recall 0.72, F1 0.64
3) **Change in NEP status since t-1**
SVM: accuracy 0.64, macro-recall 0.60, F1 0.46
4) **Change in NEP status at t+1**
LR: accuracy 0.59, macro-recall 0.63, F1 0.44

Analysis of the loadings of the most important principal components highlighted that DMOs related to walking speed, stride length and stride duration are most predictive of disease progression. The good performance of these models at predicting NEP status at t+1 emphasises the validity of real-world mobility monitoring as an approach to detecting people at risk of disease progression.

V. Conclusion

The models performed better at overall prediction of NEP status (targets 1 and 2) than at predicting when that change would happen (targets 3 and 4). The top performing models for NEP status at t+1 outperformed the top performing models for NEP status at t. These results demonstrate that novel digital biomarkers derived from IMU data have the potential to act as surrogate markers for disease progression in pwMS.

Future work could focus on improving the accuracy of prediction of disease progression in pwMS by: 1) refining feature extraction from the raw IMU signals, and 2) developing more sophisticated deep learning models that are better suited to capturing temporal patterns in the underlying data.

References

[1] Dobson R, et al. Multiple sclerosis – a review. Eur J Neurol. 2019;26:27–40.
[2] Scalfari A, et al. Smouldering-Associated Worsening in Multiple Sclerosis. Ann Neurol. 2024;96:826–845.
[3] Jakimovski D, et al. Multiple Sclerosis. Lancet. 2024;403:183–202.
[4] Del Din S, et al. Mobilise-D Technical Implementation Manual (V1.0). 2025.

[1]University of Sheffield [2]Sheffield Teaching Hospitals [3]Newcastle NIHR BRC, Newcastle University [4]University College Dublin [5]Robert Bosch Krankenhaus & Heidelberg University [6]Universitätsklinikum Schleswig-Holstein [7]IS Global [8]Harvard University [9]Uniklinikum Erlangen [10]Norges teknisk-naturvitenskapelige universitet NTNU [11]Christian Albrechts Universität zu Kiel [12]Università Vita-Salute San Raffaele