Predicting Parkinson's Disease Severity from Gait with Federated Learning: Challenges to Overcome.

Chloe Hinchliffe[1], Hugo Hiden[2], Emma Packer[1], Phillip Brown[3], Alison J Yarnell[1,3,4], Lynn Rochester[1,3,4], Lisa Alcock[1,3], Marloes Peeters[5], Silvia Del Din[1,3,4], and Paul Watson[2].

[1]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom.
[2]School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom.
[3]The Newcastle upon Tyne Hospitals NHS Foundation Trust, United Kingdom.
[4]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.
[5]School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom.

BACKGROUND: In this talk, we will describe the system being built in the TORUS project [1], in which cameras and sensors will be installed in the homes of people with Parkinson's Disease (PD) so that their health can be monitored. Traditional machine learning would require data to be transferred to a central location, but this raises serious privacy issues. We are therefore exploring an alternative, federated learning (FL), which shares model parameters instead of patient information. To this end, we will explore the use of FL to predict PD severity from real-world digital measures of gait.

METHODS: As a part of the CiC study [2], 55 people with Parkinson's were assessed using the Movement Disorder Society-Unified Parkinson's Disease Rating (MDS-UPDRS) and then wore an inertial movement unit (IMU) on the lower-back for 7 continuous days. Daily gait measures were then extracted from this data using validated algorithms [3]. These gait measures, along with patients' age, sex, and body mass index, were used to estimate the MDS-UPDRS Part III score (which is specific to mobility and ranges from 0-132) in a federated system, as well as a centralised system for comparison. For the FL system, each client was given data from a single patient. Local client models were trained, and the model parameters were aggregated by the central server, giving the global model which was then redistributed to the clients. The clients tuned the global model to the local training data and then evaluated on the local testing data. For both systems, 10-fold cross-validation was used to evaluate the model's (global model for FL) accuracy when predicating MDS-UPDRS Part III score in new patients. Two models were tested for both systems: Random Forest (RF) and fully connected Neural Networks (NN).

RESULTS: 53 patients had available gait data for analysis. The centralised models performed reasonably well, with mean absolute errors (MAEs) of RF=8.54 and NN=10.01. The MAEs of the global models were FL-RF=8.13 and FL-NN=9.27, and the mean MAEs of the local models were FL-RF=0.0 and FL-NN=4.09. However, the Pearson correlations between the (global) predicted and true scores were RF=0.141, NN=0.165, FL-RF=-0.374, FL-NN=0.09, therefore the FL-RF had a substantially poorer performance in its predictions, which is not reflected in the MAE. Notable challenges, such as increased computation cost, arose when applying FL to these data, especially for the RF. Since the local models were trained and tested with the same label, the patient's UPDRS score, the local models were severely overfit, leading to "perfect accuracy" for the local RFs but poor correlation between true and predicted scores for the global RF models.

CONCLUSIONS: This work has highlighted key challenges of using FL with individual participants at each client: RF outperformed NN for the centralised system, but FL-NN was the

best overall model. Our goal is to overcome these challenges, leading to an accurate AI system that prioritises privacy and personalisation.

REFERENCES: [1] https://torus.ac.uk/ [2] Packer et al *BMJ Open* 2023. [3] Del Din et al *J Neuroeng Rehabil* 2016.