Predicting lifestyle health behaviours using ML with EMA and fitness tracker data: Feasibility study

Hana Sediva[1], Prof. Louise Thomas[1],
Dr Tina Cartwright[2], Dr Claire Robertson[1], Dr Sanjoy K. Deb[3]

[1] School of Life Sciences, University of Westminster, UK [2] School of Social Sciences, University of Westminster, UK, [3] Nutrition, Exercise, and Health Sciences, Anglia Ruskin University, UK

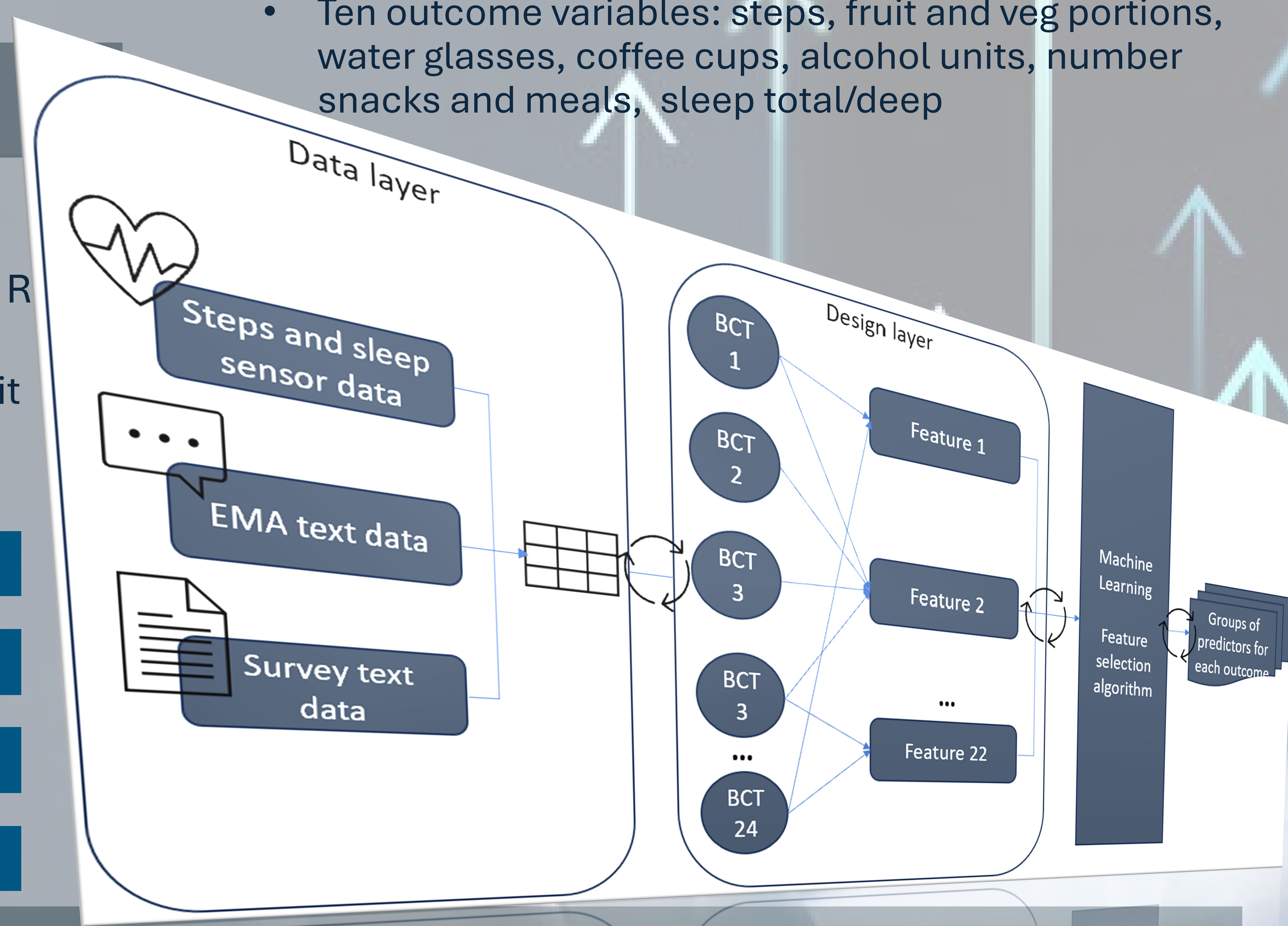
UNIVERSITY OF WESTMINSTER

Background

- Changes in complex health behaviours are idiosyncratic (differ between individuals), dynamic (fluctuate over time), and multi-factorial (influenced by many variables, such as low energy, poor sleep, social environment) [456]
- Human behaviour is complex, often influenced by many factors, and therefore difficult to predict [7]
- Effective BCIs need to predict and explain health behaviours at an individual-level [8]

Methods

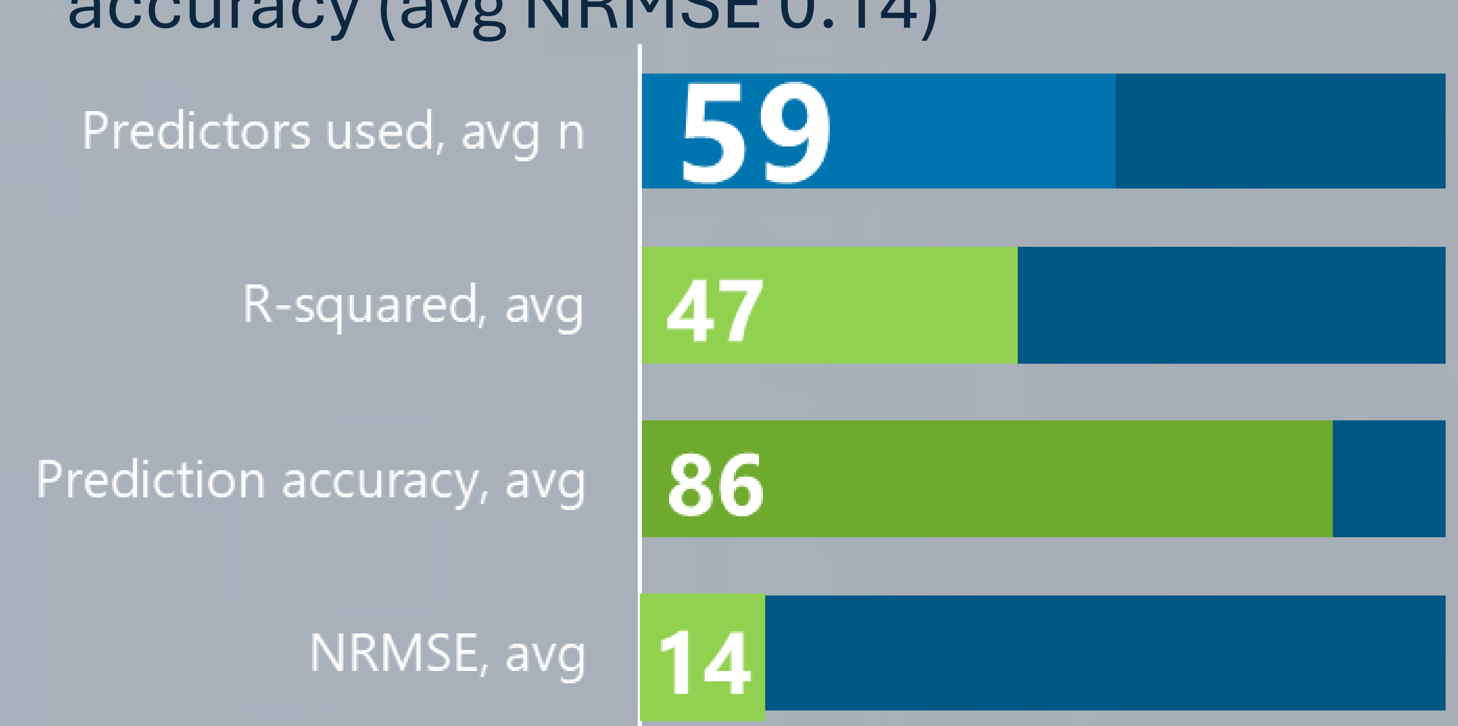
- Group-level supervised regression ML algorithms (Random Forest, Gradient Boosting Regressor) and Feature selection using Recursive Feature Elimination (RFE) algorithm
- Explanatory variables: a sample of 22 time-varying predictors (e.g., intervention data from EMA/fitness tracker) and 20 time-constant predictors (e.g., demographics from baseline survey), based on a multimodal dataset from a14-day intervention
- Ten outcome variables: steps, fruit and veg portions, water glasses, coffee cups, alcohol units, number snacks and meals, sleep total/deep



Results

✓ Feasibility: Feature selection identified groups of predictors for each outcome; Python (Gradient Boosting Regressor) and R (Random Forest) produced similar results

✓ Acceptability: Acceptable goodness-of-fit (avg R-squared 0.47) and prediction accuracy (avg NRMSE 0.14)



Conclusion

1. Using ML to identify most relevant behavioural predictors, linked to BCTs, has the potential to create more adaptive, dynamic, and personalised lifestyle health interventions in future studies.
2. Goal setting (e.g., morning exercise plan, steps, veg portions), and counselling (e.g., problem solving, action planning) were most frequently selected group-level predictors. Time-constant predictors, including ethnicity and cultural background should be considered to improve personalisation
3. Human coach interaction had a significant effect on increasing engagement and reviewing educational content on diet, therefore, human-supported or hybrid (VR) interventions should be considered in future studies

Limitations:

- Small dataset (of 137 records) may lead to model overfitting
- The groups of predictors are only examples and should not be used as the most effective groups of predictors
- The selected ML regressors are only used for testing purposes and are not necessarily the best options

References:

4. Arend, A.-K. et al. (2023). Toward Individualized Prediction of Binge-Eating Episodes Based on Ecological Momentary Assessment Data: Item Development and Pilot Study in Patients With Bulimia Nervosa and Binge-Eating Disorder. *JMIR medical informatics*, 11 (1), e41513.
5. Chevance, G., Perski, O. and Hekler, E.B. (2021). Innovative methods for observing and changing complex health behaviors: four propositions. *Translational Behavioral Medicine*, 11 (2), 676–685. Available from https://doi.org/10.1093/TBM/IBAA026
6. Perski, O. et al. (2023). Classification of Lapses in Smokers Attempting to Stop: A Supervised Machine Learning Approach Using Data From a Popular Smoking Cessation Smartphone App. *Nicotine & Tobacco Research*, 25 (7), 1330–1339.
7. Ozili, P.K. (2023). *The Acceptable R-square in Empirical Modelling for Social Science Research*.
8. Hekler, E.B. et al. (2016). Advancing Models and Theories for Digital Behavior Change Interventions. *American Journal of Preventive Medicine*, 51 (5), 825–832.