Reproducible machine learning of wearable sensor data for the early detection of cardiovascular disease

Aiden Doherty[1,2,3]

[1] Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, U.K.

[2] Nuffield Department of Population Health, University of Oxford, Oxford, U.K.

[3] National Institute of Health Research Oxford Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, U.K.

Cardiovascular disease (CVD) prevention strategies include the typical use of risk stratification and prediction tools to target preventive interventions for people at higher risk of CVD and recommendations on modifiable CVD risk factors such as levels of physical activity. However, most people with CVD are identified too late and over 50% of major cardiac events are in patients who were not classified as high-risk. In addition, a reliance on crude self-reported questionnaires could mean that CVD behavioural risk factors such as physical activity and sleep duration are more important than previously thought.

Wearable sensors such as wrist-worn activity trackers (accelerometers) have the potential to continuously, noninvasively, and painlessly measure CVD risk factors in patients' everyday lives. For example, my group has worked closely with UK Biobank to measure physical activity status in 103,712 participants who agreed to wear a wrist-worn device for seven days. These measurements are now actively used by health researchers worldwide to demonstrate associations between physical activity and CVD. Machine learning methods can help maximise the utility of data from wearable sensors. However, there is a broad concern around the lack of reproducibility of machine learning models in health data science. It is critical to carefully consider how to promote robust machine learning findings and reject irreproducible ones, to ensure credibility and trustworthiness.

In this talk I will share my groups work on reproducible machine learning of wearable sensor data for the early detection of cardiovascular disease. This will include methods to identify physical activity behaviours in a free-living validation dataset of ~150 adults. I will then illustrate the genetic architecture of these measurements. I will also show that these measurements have a clear dose-response association with incident CVD outcomes after ~5 years of follow-up. Finally, I will discuss the opportunities for wearable sensors to advance the prevention of CVD.