**Machine-learning clustering of individuals at risk of dementia using electronic healthcare records**

**Research Context**

As the global population ages, dementia has emerged as a significant health and societal challenge, with few treatments offering limited relief. New drug development for dementia has been very difficult due to insufficient understanding of this complex disease. However, some risk factors of dementia have been established, including hypertension, diabetes, and depression. Our project, Drug Repurposing and targetIng for dementia preVENtion (DRIVEN) is to evaluate if existing medications that treat potential risk factors of dementia (e.g. antihypertensives, antidiabetics, and antidepressants) could be repurposed for dementia prevention.

One important consideration is that individuals at risk of dementia often present different clinical characteristics, which could lead to varying responses to medications and different prognoses of dementia. It would be very useful to identify those who would benefit the most from the medications to inform clinicians and policy makers about the target population for dementia preventive treatment. In the first instance, this requires understanding of the clinical characteristics of individuals at risk of dementia, and how they could be classified into clinical groups according to these characteristics. The effects of the repurposed medications for dementia prevention could then be further assessed within each group to identify the optimal target populations for dementia preventive treatment.

**Internship Project**

This internship project will focus on developing a machine-learning algorithm to classify and characterise groups of individuals at risk of dementia based on their clinical features. This will facilitate our broader DRIVEN work to identify individuals for targeted preventive treatment. The algorithm will be developed by applying unsupervised or semi-supervised machine-learning clustering approaches in the ORCHARD databases, which contain electronic health records of over 30,000 hospital patients with baseline cognitive screening assessments and follow-up dementia outcomes. The algorithm will be generalised and applied to a wider population from the CPRD databases, which include electronic health records of millions of patients in primary care.

In this internship, you will have guidance in both medical research and machine-learning techniques. You will learn the key steps involved in applying cutting-edge machine-learning to solve a real-world problem, develop practical skills of the popular programming languages and gain experience in handling large-scale healthcare data. You will be working closely with multiple research groups within the University of Oxford with opportunities to attend online or in-person training resources (e.g. DataCamp) as well as potential opportunities to attend workshops, seminars, and conferences.

The applicant will benefit from some knowledge of quantitative subjects (e.g. computer sciences, engineering, statistics, mathematics, epidemiology, bioinformatics) and programming. We especially encourage students with a strong interest in applied machine-learning in medical research, but all levels of relevant knowledge and skills in this area are welcome. This internship is envisaged to be 1 day a week for 12 months, but there is flexibility in this timescale so applicants could fit this around their schedule, for example it could be done 2 days a week over 6 months.