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

**Context**

Nearly a million people in the UK are living with dementia. Large numbers of people living with dementia is a growing problem in health and social care as people living with dementia have a poor quality of life and worse health outcomes. Initial symptoms of dementia are memory loss or forgetfulness, and this is often when the initial consultation happens with primary care physicians. The time taken to progress from initial symptoms to confirmed dementia diagnosis is very variable and there could be many factors that affect the rate at which people progress. We are trying to understand how the progression to dementia happens in different people according to factors such as, gender, age group, ethnicity, and presence of other long-term health conditions.

**Research context**

The journey of an individual towards dementia, and the associated factors, can help to detect the stage where intervention could be taken to reverse or slow the process. We will examine if, and how, routine tests can be related to dementia and what happens next. Understanding how changes in routinely collected information on factors such as blood pressure, and blood sugar levels, could predict dementia much earlier and allow intervention at an optimal stage without relying on expensive tests. To start to answer this, we will use the largest primary care record available in the UK and develop some advanced machine learning methods. We will also group people based on outcomes such as death, and hospital admission using advanced statistical methods.

**Research project**

This internship project will involve learning about management of big data sets and analysis using advanced statistical machine learning and artificial intelligence methods. The project involves one of the largest sets of real-world health care data in the UK covering nearly 60 million people. The Intern will have access to short training courses and hands on training to learn multiple skills such as data science, statistics, coding, and epidemiological research using real world data.

This internship is expected to be 1.5 – 2 days a week over 12 months.