

## **Industrial Maths study groups – first problem announced for the AI and Health event**

The first in a series of problems for KTN's Industrial Maths in AI and Health event has been announced and researchers are encouraged to apply to take part.

KTN's Industrial Maths Knowledge Transfer Manager, [Matt Butchers](#), regularly runs study groups looking at problems in different sectors. There are several forthcoming groups, each of which will consider four-five different problems over a three-day event. The groups will focus on AI and Health, Clean and Sustainable Growth and Uncertainty Management.

The first problem to be tackled at the AI and Health study group has been announced. The organisers, KTN, alongside the Universities of Cardiff and Manchester, are looking for researchers to work on the following conundrum:

***Secure Machine Learning for Rare Disease Prediction presented by Mendelian.***

***Challenge context:*** According to the EU, a rare disease affects fewer than 5 in 10,000 people. Collectively, they affect a large portion of the population and cause a significant burden on our healthcare systems. There are an estimated 3.5 million rare disease patients in the UK (that's more than all cancer patients in the UK)[1]. On average, these patients wait 5.6 years for a diagnosis and are referred to 7.3 different doctors [2].

*The problem is that rare diseases are difficult to diagnose. They are numerous and varied, making it difficult for doctors to recognise them on demand. There have been several initiatives to organise medical knowledge around rare disease like Orphanet. To improve diagnosis, Genome England has allocated half of the 100,000 genome project to tackle rare disease [3].*

*Mendelian has spent several years creating tools for clinicians, geneticists and healthcare systems who manage rare diseases. Our clinical search engine has had much success with doctors seeking the latest relevant knowledge for their troublesome diagnoses. We have ongoing studies showing how our technology can improve rare disease screening and clinical decision making.*

*In academia, there has been machine learning research on predicting rare disease in electronic health records [4,5]. These studies have shown that machine learning methods are able to predict diagnoses that expert clinicians would make. Unfortunately, it is often infeasible to find large enough datasets to sufficiently train classifiers for diseases with such a low prevalence. In practice, it would require access to data held by several institutions that is very sensitive in nature.*

*We are interested in machine learning methods that are effective and secure on siloed data. There has been a lot of interest in Federated Learning and Zero-knowledge computation to tackle issues with disparate, sensitive data [6,7]. Our hope is that methods like these could make machine learning a feasible for rare disease prediction while preserving patient privacy. This research raises questions around the level of privacy that can be guaranteed; if there are methods to allow hospitals to moderate the data leaving their system; and which restrictions would this have (if any) on predictive performance.*

**Resources available for challenge:** *Mendelian can provide a patient data for ideation and to validate proofs of concept. These datasets are either limited, anonymised, public domain patient records or synthetic data produced from the latest rare disease statistics. For well-formed solutions, we would be able to test them on larger, secured patient record databases to feedback summary statistics and performance metrics.*

## **References:**

1. What is a rare disease? Rare disease UK, viewed 4 December 2018, <https://www.raredisease.org.uk/what-is-a-rare-disease/>
2. Engel, P.A. & Bagal, S & Broback, M & Boice, N. (2013). Physician and patient perceptions regarding physician training in rare diseases: The need for stronger educational initiatives for physicians. *J Rare Dis.* 1. 1-15. <http://www.journalofraredisorders.com/pub/IssuePDFs/Engel.pdf>
3. 100,000 Genome Project Rare Disease Programme, viewed 4 December 2018 <https://www.rcplondon.ac.uk/file/8891/download?token=5ru8ybmi>
4. Tremblay, Michael & Colbaugh, Richard & Glass, Kristin & Rudolf, Christopher. (2018). Robust ensemble learning to identify rare disease patients from electronic health records. 10.13140/RG.2.2.19677.92645. Available at: [https://www.researchgate.net/publication/326260671\\_Robust\\_ensemble\\_learning\\_to\\_identify\\_rare\\_disease\\_patients\\_from\\_electronic\\_health\\_records](https://www.researchgate.net/publication/326260671_Robust_ensemble_learning_to_identify_rare_disease_patients_from_electronic_health_records)

5. Garg et al. (2016). A Bootstrap Machine Learning Approach to Identify Rare Disease Patients from Electronic Health Records. Available at: <https://arxiv.org/abs/1609.01586>
6. Federated Machine Learning. Towardsdatascience.com, viewed 4 December 2018, <https://towardsdatascience.com/federated-machine-learning-c99dd5dec201>
7. Riaz et al. (2018). Chameleon: A Hybrid Secure Computation Framework for Machine Learning Applications. Available at: <https://arxiv.org/abs/1801.03239>

In recent years, there has been a huge surge in interest in artificial intelligence (AI) for health and care. In a recent speech by the Prime Minister, it was asserted that: “the United Kingdom will use data, artificial intelligence and innovation to transform the prevention, early diagnosis and treatment of diseases like cancer, diabetes, heart disease and dementia by 2010”.

There are huge commercial opportunities in public health, vaccination, medicine discovery and manufacture, mental health services, medical technology for monitoring lifestyle and behaviour.

A key component to developing a successful and booming AI in health and care eco-system is providing tangible examples of the successful application of AI detailing the techniques, successes, and limitations on real data.

This study group will bring researchers from across the UK together with a number of industrial organisations (large, governmental, SMEs) to tackle some of the most interesting challenges in this sector.

This is the first of several problems that will be discussed at two three-day study groups, firstly in Cardiff from 22-24 May and secondly in Manchester from 26-28 June.

If you are a researcher working in a UK university who would like to work on this problem, please register [here](#).