MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Singapore-NZ Data Science Workshop: Summary of Health Theme discussion 30-31 October, 2019

Participants:

Sebastian Maurer-Stroh	sebastianms@bii.a-star.edu.sg	Bioinformatics Institute
Gavin James Smith	gavin.smith@duke-nus.edu.sg	Duke NUS Medical School
Brian Premchand	brian_premchand@i2r.a-star.edu.sg	Institute for Infocomm Research
Mong-Li Lee	leeml@comp.nus.edu.sg	Institute of Data Science, School of Computing
Wynne Hsu	whsu@comp.nus.edu.sg	
Donia Macartney-Coxson	Donia.Macartney-Coxson@esr.cri.nz	Institute of Environmental Science and Research
Liu Yong	liuyong@ihpc.a-star.edu.sg	Institute of High Performance Computing
Arijit Khan	arijit.khan@ntu.edu.sg	Nanyang Technological University
Zinovi Rabinovich	zinovi@ntu.edu.sg	
Alex Cook	ephcar@nus.edu.sg	National University of Singapore
CHUI Chee-Kong	mpecck@nus.edu.sg	
Matthew Chua	mattchua@nus.edu.sg	
Edmond Zhang	edmond.zhang@orionhealth.com	Orion Health
Ee-Peng Lim	eplim@smu.edu.sg	Singapore Management University
Alexei Drummond	alexei@cs.auckland.ac.nz	University of Auckland
Gillian Dobbie	g.dobbie@auckland.ac.nz	
Colin Simpson	colin.simpson@vuw.ac.nz	Victoria University of Wellington

This report aims to capture the ideas that were generated during the 2-day Singapore-NZ Data Science Workshop in the Health theme. On Day 1 we collected high level ideas and on day 2, we described possible projects. There were rich discussions, and the findings have been summarised here.

The ideas of the first day were grouped into clusters:

- non communicable diseases
- infectious diseases and surveillance
- methodology and trans-country studies
- interpretation/explainability
- data sets
- privacy

Some of the ideas in each cluster included:

Non communicable diseases

- metabolic disease like diabetes studies, diet/ethnicity/lifestyle
- dementia, investigate factors in each country
- push management of chronic diseases from hospital to home
- effects of SNPs in inter-individual variability in response to disease and response to drugs and environment

Infectious diseases and surveillance

- infectious diseases and surveillance and outbreak analysis
- real-time infectious disease analysis platform (geospatial+surveillance+medical records+genome data)
- modelling transmission with infectious disease hub and possible interventions

Methodology and trans-country studies

- improving prediction by combining first principles and machine learning
- GWAS vs/+ machine learning risk prediction in non-European populations (genome and medical record data)
- localising prediction models for a multi-cultural population
- longitudinal studies
- replication/validation - interoperability, ontologies, data sovereignty
- AI for triaging
- multi-model AI for disease detection and prediction
- federated learning for cross country data sharing

Interpretation/explainability

- explainable AI for knowledge transfer particularly end user empowerment
- explainable AI for medical diagnosis
- explanatory models vs/+ explainable machine learning models in complex diseases/genomics
- bias/transparency/fairness/explainability/accountability with medical data
- bias/transparency - understanding how to measure it and its influence, understand how machine learning algorithms are impacting subpopulations

Data sets

- build individual healthcare data sets for Singapore and NZ, such as MIMIC III (Medical Information Mart for Intensive Care- a database for information relating to patients admitted to intensive care)
- survey of available health data sets
- knowledge base graph construction with hospital data and conversational AI

Privacy

- privacy preserving AI on healthcare data

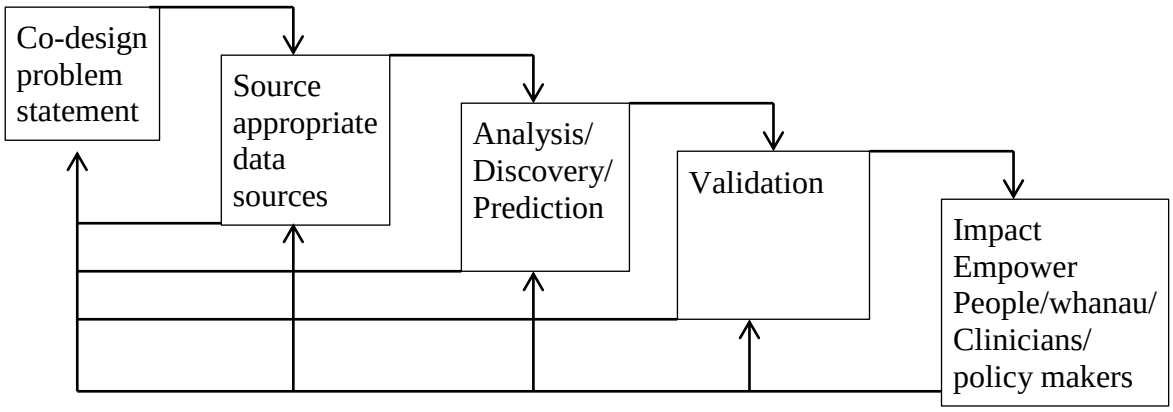
With respect to data sets, we recognised the following issues:

- incentive to share (eg. primary data not well linked up), identifying right level for anonymisation
- creation, management and capacity – infrastructure needs
- new data sources - social determinants, public transport, meteorological/pollution
- multi-modal, imaging, free text, spatial temporal, timely

A couple of umbrella concerns were:

- public and patient involvement
- a need for learning health systems (actionable data)
- data sovereignty

The following diagram outlines a general data science pipeline:



We described two potential projects, as examples, and then collected a list of other project ideas.

Example Project: Type 2 Diabetes

Aim: early detection of type 2 diabetes that could lead to severe outcomes (and possibly include interventions)

Data science stretch:

- integrate diverse data sources (e.g., Genetics Aotearoa data)
- look at multimorbidity (instigator of this work is Prof. Stewart Mercer)
- use explainable neural nets (graphNN)
- detecting and countering bias
- bring together traditional and non-traditional techniques to get better understanding of techniques and results

Suggestions for possible partners:

- data providers
- genomic information owners
- Thomas Lumley
- SERI (Singapore Eye Research Institute)
- Singapore Health Promotion Board
- National Kidney Foundation
- Duke/NUS

Example Project: Infectious diseases in humans

Aim: study global circulation dynamics, vaccine effectiveness, antimicrobial resistance, improved epidemic prediction, simulate spread and adjust parameters, early detection from medical records, model molecular evolution

Data science stretch:

- Distinguish between different viruses, and sources of viruses
- Compare and combine black box machine learning models and molecular evolution models (validating/explaining)
- Integrating data streams, data stream analysis
- Systematic differences in social media, e.g., cultural

Suggestions for possible partners:

- Epidemiologists
- Centre for infectious diseases
- Centre for Infectious Disease Epidemiology and Research (CIDER)
- ESR
- Sue Huang
- Michael Baker

Other project ideas:

- alzheimers
- dementia
- depression
- mental health
- autism
- healthy ageing
- IoT (Internet of Things) in health
- Entity recognition in health
- E-consent
- Privacy and computation
- Understanding cultural differences around social license