A Case Study in Data Science and AI
Predicting Organ Failure in Hospitals

Shakir Mohamed
@shakir_za
shakir@deepmind.com

#DSRD19
Many areas for Machine Learning and Digital Platforms to play a role.
Triple Aims

1. Better clinical outcomes
2. Enhance patient and clinician experience
3. Reduce costs

Systemic challenges

- >50% of healthcare not evidence based
- Staff burnout rates on the rise
- Care continues to be episodic vs integrated
- Intractable increases in healthcare costs
- Failure to deliver shared decision making for patients
- Unwarranted variation exists across healthcare delivery
- >10% of patients experience harm in hospitals
- Focus and on illness at the expense of prevention
• Millions of people die every year from diseases that could be prevented with earlier detection.
• Worked with a hospital partner to look at AI for predicting patient deterioration.
• **Acute kidney injury (AKI)**, a condition where a patient’s kidney suddenly stops working properly. Affecting up to 1 in 5 hospitalised patients in UK and US.

Data from these processes are captured within an electronic health record.
Duplicate report(s) will be sent to the following copy-to physician(s):
FAIRFAX FAMILY PRAC. CTR., STE 400, 3650 JOSEPH SIEWICK DR, FAIRFAX VA

**CBC (INCLUDES DIFF-PLT)**

<table>
<thead>
<tr>
<th>TESTS</th>
<th>RESULT</th>
<th>FLAG</th>
<th>UNITS</th>
<th>REF</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHITE BLOOD CELL COUNT</td>
<td>3.7</td>
<td>L</td>
<td>Thousand/ul</td>
<td></td>
</tr>
<tr>
<td>RED BLOOD CELL COUNT</td>
<td>3.39</td>
<td>L</td>
<td>Million/ulL</td>
<td></td>
</tr>
<tr>
<td>HEMOGLOBIN</td>
<td>10.7</td>
<td>L</td>
<td>g/dL</td>
<td></td>
</tr>
<tr>
<td>HEMATOCRIT</td>
<td>30.3</td>
<td>L</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>MCV</td>
<td>89</td>
<td></td>
<td>fl</td>
<td></td>
</tr>
<tr>
<td>MCH</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Today: Wednesday 8/27/2014 12:45:36 PM  Press F1 for Help  Status: Logged in user: Jane
Characteristics
• Unstructured
• Noisy
• Recorded differently

DS/AI Interactions:
• SWE and architects
• Security, privacy, law
• Clinical requirements
## Data and Summarisation

### Data from a large hospital partner

<table>
<thead>
<tr>
<th>Patients</th>
<th>Training</th>
<th>Validation</th>
<th>Calibration</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unique patients</strong></td>
<td>562,507</td>
<td>35,277</td>
<td>35,317</td>
<td>70,681</td>
</tr>
<tr>
<td><strong>Average age</strong></td>
<td>62.4</td>
<td>62.5</td>
<td>62.4</td>
<td>62.3</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>106,299 (18.9%)</td>
<td>6,544 (18.6%)</td>
<td>6,675 (18.6%)</td>
<td>13,183 (18.7%)</td>
</tr>
<tr>
<td>Other</td>
<td>456,208 (81.1%)</td>
<td>28,733 (81.4%)</td>
<td>28,642 (81.4%)</td>
<td>57,496 (81.3%)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>35,855 (6.4%)</td>
<td>2,300 (6.5%)</td>
<td>2,252 (6.4%)</td>
<td>4,519 (6.4%)</td>
</tr>
<tr>
<td>Male</td>
<td>526,652 (93.6%)</td>
<td>32,977 (93.5%)</td>
<td>33,065 (93.6%)</td>
<td>66,162 (93.6%)</td>
</tr>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56,958 (10.1%)</td>
<td>3,599 (10.2%)</td>
<td>3,702 (10.5%)</td>
<td>7,093 (10.0%)</td>
<td></td>
</tr>
<tr>
<td>22,284 (4.0%)</td>
<td>1,367 (3.9%)</td>
<td>1,384 (3.9%)</td>
<td>2,784 (3.9%)</td>
<td></td>
</tr>
<tr>
<td><strong>Renal replacement therapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22,284 (4.0%)</td>
<td>1,367 (3.9%)</td>
<td>1,384 (3.9%)</td>
<td>2,784 (3.9%)</td>
<td></td>
</tr>
</tbody>
</table>
Data Summarisation

Characteristics
• Sequential representation
• Sparse
• Missing data
• Included and excluded
• Handling time, alignment

DS/Al Interactions:
• Important research questions; arise from practical considerations.
• Where do labels come from?
• What predictions and metrics are important to clinicians.

<table>
<thead>
<tr>
<th>Admissions within a five year period</th>
<th>Summary of dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data center sites</td>
<td>130***</td>
</tr>
<tr>
<td>Unique admissions</td>
<td>2,004,217</td>
</tr>
<tr>
<td>Average</td>
<td>3.6</td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
</tr>
<tr>
<td>Duration (days)</td>
<td>9.6</td>
</tr>
<tr>
<td>Average</td>
<td>3.2</td>
</tr>
<tr>
<td>ICU admissions</td>
<td>214,644 (10.7%)</td>
</tr>
<tr>
<td>Medical admissions</td>
<td>971,527 (48.5%)</td>
</tr>
<tr>
<td>Surgical admissions</td>
<td>354,008 (17.7%)</td>
</tr>
<tr>
<td>No creatinine measured</td>
<td>408,927 (20.4%)</td>
</tr>
<tr>
<td>Chronic Kidney Disease</td>
<td>746,692 (37.3%)</td>
</tr>
<tr>
<td>Stage 1**</td>
<td>8,409 (0.4%)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>429,990 (21.5%)</td>
</tr>
<tr>
<td>Stage 3A</td>
<td>156,720 (7.8%)</td>
</tr>
<tr>
<td>Stage 3B</td>
<td>77,801 (3.9%)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>31,646 (1.6%)</td>
</tr>
<tr>
<td>Stage 5</td>
<td>50,535 (2.5%)</td>
</tr>
<tr>
<td>AKI present</td>
<td>267,396 (13.3%)</td>
</tr>
<tr>
<td>Any AKI</td>
<td>207,441 (10.4%)</td>
</tr>
<tr>
<td>Stage 1</td>
<td>207,441 (10.4%)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>43,446 (2.2%)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>66,734 (3.3%)</td>
</tr>
</tbody>
</table>
Future Prediction of AKI

Useful predictions are those that are accurate and continuously updated, given with sufficient time to act, provide context for decision.

Model on 700k features. Make predictions up to 48hrs ahead.
Models

- Focus on strong baselines that were the current state of the art.
  - Gradient Boosted Trees
  - Logistic regression
- New models using Deep Learning
  - Non-linear models and interactions
  - Continuous integration of information as they are received
- Calibration, uncertainty
a) Patient creatinine measurements during admission

AKI occurs 48 hours after the model first generated a positive prediction.

b) Model predictions for any AKI within 48 hours

Predictions above the risk threshold indicate a positive prediction of AKI within 48 hours. Model predicts AKI within 48 hours.

48 hours

Time since admission (days)

0 2 4 6 8 10

Patient risk of AKI within 48 hours

0.0 0.2 0.4 0.6 0.8 1.0

Model prediction
0.2 risk threshold
Prediction uncertainty

Future prediction from a fixed point (hours)

6 12 18 24 36 48 60 72

Serum creatinine (μmol/L)

0 80 160 240 220

Predicted maximum serum creatinine
Maximum recorded serum creatinine
Measurement of serum creatinine at time 0

Future prediction from a fixed point (hours)

5 10 15 20 25 30 35 40 45

Serum urea nitrogen (mg/dL)

5 10 15 20 25 30 35 40 45

Predicted maximum serum urea nitrogen
Maximum recorded serum urea nitrogen
Measurement of serum urea nitrogen at time 0

Future prediction from a fixed point (hours)

2.5 3.0 3.5 4.0 4.5 5.0

Serum potassium (mEq/L)

2.5 3.0 3.5 4.0 4.5 5.0

Predicted maximum serum potassium
Maximum recorded serum potassium
Measurement of serum potassium at time 0
Summary:
• Make predictions of AKI up to 48hr ahead.
• Provide a window for meaningful action.
• For the most severe cases, can detect up to 90% of cases.

Further considerations and limitations:
• Early or late predictions and alerting fatigue
• Generalisation needed to wider steps of hospitals, patient populations.
• Only a retrospective study.
• Need prospective studies to evaluate real clinical-use.
Many Other Sources of Questions, Partners and Data
Statistical Operations

- Estimation and Learning
- Modelling
- Data Enumeration
- Comparison
- Inference
- Hypothesis Testing
- Experimental Design
Statistical Operations

Inference

What we can know about our data

Decision-making

What we can do with our data.
Centrality of Inference

The core questions of AI will be those of probabilistic inference.

Artificial Intelligence will be the refined instantiation of these statistical operations.
# Principals to Products

## Applications
- Assistive Technology
- Advancing Science
- Climate and Energy
- Healthcare
- Fairness and Safety
- Autonomous systems

## Reasoning
- Planning
- Explanation
- Rapid Learning
- World Simulation
- Objects and Relations

## Information
- Uncertainty
- Information Gain
- Causality
- Prediction

## Principles
- Probability Theory
- Bayesian Analysis
- Hypothesis Testing
- Estimation Theory
- Asymptotics
Neutrality and Universality

Neutrality Traps

• **The Portability Trap:** solutions designed for one social context may be inaccurate / do harm when applied to a different context.

• **The Formalism Trap:** Failure to account for the full meaning of social concepts such as fairness, and think they can be resolved through mathematical formalisms.

• **The Ripple Effect Trap:** Inserting technology into an existing social system changes the behaviours and embedded values of the pre-existing system.

• **The Solutionism Trap:** Failure to recognise the possibility that the best solution to a problem may not involve technology.

Universality

‘A mono-cultural view of ethics conceives itself as the only valid one. In order to **avoid this kind of ethical chauvinism and colonialism** it is necessary that transcultural ethics arise from an intercultural dialogue instead of thinking of itself as universal without noticing its own cultural bias.’ Capurro, 2004
The Global Goals for Sustainable Development

1. No Poverty
2. Zero Hunger
3. Good Health and Well-Being
4. Quality Education
5. Gender Equality
6. Clean Water and Sanitation
7. Affordable and Clean Energy
8. Decent Work and Economic Growth
9. Industry, Innovation and Infrastructure
10. Reduced Inequalities
11. Sustainable Cities and Communities
12. Responsible Consumption and Production
13. Climate Action
14. Life Below Water
15. Life on Land
16. Peace and Justice Strong Institutions
17. Partnerships for the Goals
Predicting Organ Failure in Hospitals

A Case Study in Data Science and AI

Predicting Organ Failure in Hospitals