Reinventing Healthcare with Big Data & Machine Learning

Farah Shamout
PhD Student in Engineering Science
Computational Health Informatics Lab
University of Oxford
Data in Healthcare
Research Focus

1. Personalized patient time-series modeling due to the added advantage of **massive scale** and the possibility of **patient-specific analysis**.

2. **Novelty detection of clinical deterioration** through **fully-predictive means**, rather than merely identifying deterioration as it begins.
Motivation: Detecting Clinical Deterioration

1. One of the main contributory factors is **inefficient ward management**.
2. **High costs** of preventable adverse events (Prolonged hospital stays, litigation, staff time, burden on patients, and broader economic consequences)
3. **Massive scale of available data**, acquired from the HAVEN Database.
4. Most existing models **merely identify deterioration as it begins**.

Detection Approaches

- **Traditional Methods**
- **Machine Learning Methods**
Clinical Deterioration refers to the worsening of a patient’s condition on hospital wards, and is assessed by medical staff through routine observations and protocols.

Preventable adverse events are avoidable injuries or complications where there was enough knowledge and accepted practices to have avoided the event [15].
Electronic Health Records (EHRs) - HAVEN Database

Patient Cohort Size

HAVEN v14

HAVEN v14
1. **Definition:** “Longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting” - HIMSS

2. **Uses:** Extraction of patient cohorts (phenotypes), early warning scores, machine learning

3. **Challenges:**
   a. Complexity
   b. Completeness
   c. Correctness
   d. Currency

4. **Data Preparation for Analysis:**
   a. Raw data
   b. Derived Parameters
   c. Statistical Measures
   d. Time Series Features
Snapshot of a Patient Admission

ICU Admission
**EHRs Workflow**

**Step 1: Extract Patient Cohort**
1. Elective vs. Emergency Admissions
2. Surgical vs. Non-surgical patients
3. Clinical Phenotype

**Step 2: Data Pre-processing**
1. Physiologically implausible values
2. Sparsity & missingness
3. Varying lengths of stay
4. Fusing different types of data
5. Effect of interventions
6. Multiple consecutive outcomes

**Step 3: Data Processing**
Traditional Methods vs. Machine Learning methods
Traditional Methods: Early Warning Scores

- **Common Objective:** secure timely and effective responses for patients with deteriorating medical conditions, rather than to predict a particular outcome (RCP, 2012).
- **Methodology:** Mainly *heuristically-developed*, while some are *statistical*.
- **Examples:** ViEWS, NEWS, CEWS, MEWS, MCEWS, etc…

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aids medical staff</td>
<td>Lack of generalizability across hospitals</td>
</tr>
<tr>
<td>Easy to understand and implement</td>
<td>Limited input variables</td>
</tr>
<tr>
<td>Good performance</td>
<td>Assumes homogeneity across patients</td>
</tr>
</tbody>
</table>
Machine Learning (ML) is an application of Artificial Intelligence (AI) that allows computers to learn automatically from experience (i.e. retrospective data). Examples: classification of handwritten digits, self-driving cars, gaming such as chess, ultrasound images segmentation, predict housing prices, etc…
1. **Personalized Early Warning Score**

<table>
<thead>
<tr>
<th>Vital Sign</th>
<th>Reported Changes with Age/Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate (HR)</td>
<td>Increases</td>
</tr>
<tr>
<td>Systolic Blood Pressure (SBP)</td>
<td>Gradually increases, with a higher incidence amongst females</td>
</tr>
<tr>
<td>Diastolic Blood Pressure (DBP)</td>
<td>Increases until the fifth decade and then slowly decreases</td>
</tr>
<tr>
<td>Respiratory Rate (RR)</td>
<td>Increases</td>
</tr>
<tr>
<td>Temperature (TEMP)</td>
<td>Core body temperature decreases</td>
</tr>
</tbody>
</table>

Motivates the development of an Age- and Sex- Based Early Warning Score to alert for adverse events within 24 hours to event. Adverse event defined as cardiac arrest, ICU admission or mortality.
2. Predictive Inference

1. Time Series Modelling

2. Extraction of Features

Gaussian Processes

Deep Learning Model

Modelled Vital Signs
3. Stroke prediction

1. **Strokes**, or cerebrovascular accidents, are the **second leading cause of death worldwide** according to the World Health Organization.

2. The **burden of stroke** due to illness, disability and early death is expected to **double worldwide during the next 15 years**.

3. **Current objectives:**
   a. Predict the risk of having a stroke for patients admitted into hospital
   b. Distinguish between ischaemic and haemorrhagic strokes
Who, what and where?

Hospital

Nursing

Home

Pharma

Public Spaces

Primary Care

Sex

Age

Diseases

Lifestyle
1. **Data Collection-driven Solutions**

*System for Electronic Notification & Documentation*

(http://www.send-system.co.uk/)

**Depression management using self-reported data**

(http://www.thelancet.com/pdfs/journals/lanpsy/PIIS2215-0366(15)00471-X.pdf)
2. Diagnostic Tools

Diagnosis of sleep apnea amongst children


Early Warning Score for Pregnant Women

3. Prevention

Intelligent Hand Hygiene using Computer Vision

(https://aicare.stanford.edu/projects/hand_hygiene/)

Real-time fall detection for elderly patients

(https://arxiv.org/abs/1711.11200)
Translation of research into clinical settings

1. Generalizability of models across different clinical settings

2. Availability of an infrastructure to support new technologies

3. Designing and delivering solutions alongside ‘Human Factors’ approaches

4. Creating solutions to real and relevant problems!
Thank you for listening!

Farah Shamout
PhD Student in Engineering Science
Computational Health Informatics Lab
University of Oxford

farah.shamout@eng.ox.ac.uk