Welcome to the ISPOR Signal Series Episode

Larger, Deeper and in Real Time: Applications of Machine Learning and Natural Language Processing on Electronic Health Records

#ISPORSignal
Exploring what will shape healthcare decision making over the next decade...
Larger, Deeper and in Real Time: Applications of Machine Learning and Natural Language Processing on Electronic Health Records
Larger, Deeper, and in Real Time:

Applications of Machine Learning and Natural Language Processing on Electronic Health Records to Learn from the Patient Journey at Scale

Discussion Leader: Joe Vandigo, MBA, PhD

Discussants: Selen Bozkurt, PhD, MS
Katherine Tan, PhD
Ravi Parikh, MD, MPP
RWD is not always Patient Experience Data
(But it can be!)
Improved mapping of the patient experience has implications for RWD study designs.

**First interaction with healthcare system**

**Target Population**  
**Hypothesis Generating**  
**Study Period**

**Treatments and side effects**

**Exposure**  
**Potential Outcomes**  
**Covariates & Confounders**

Themes for today’s panel

- **Scale**
  Increasing stratification cohort sizes

- **Speed**
  Keeping up with standard of care

- **Depth**
  Improving representation of underserved populations

**Human and AI Collaboration**
Key Questions

What is the most exciting opportunity in capturing the patient journey – scale, speed, or depth?

How and when can patients and other stakeholders engage in ML/NLP processes?
Panelists

Joe Vandigo, MBA, PHD
Moderator
Applied Patient Experience

Selen Bozkurt, PhD, MS
Discussant
Stanford University

Katherine Tan, PhD
Discussant
Flatiron Health

Ravi Parikh, MD, MPP
Discussant
University of Pennsylvania
Unlocking the Power of Electronic Health Records with NLP/ML

Selen Bozkurt, PhD, MS
Senior Research/Data Scientist
Stanford University School of Medicine (Biomedical Informatics)
VA Palo Alto, Center for Innovation to Implementation
Incoming Assistant Professor, Emory University, Faculty of Medicine (Biomedical Informatics)
Disclosures

Academic consultancy services to F2IL, Flatiron Health, an independent subsidiary of the Roche Group.
Biomedical Informatics

Computer Science

Statistics

BMI

Medicine
Patient journey through EHR documentation

Radiology, Pathology Images and Reports

Labs, Meds, Progress Notes

Patient emails, messages, surveys

Longitudinally & Multi Specialty
Learning from (clinical) text

Text Input from EHR

Language Models

Numeric representations of text

Summary, Translation

Feature 1
Feature 2
...
Feature n

Features for predictors or classifiers
Learning from (clinical) text

Text Input from EHR

Language Models

Numeric representations of text

Summary, Translation

Feature 1
Feature 2
...
Feature n

Features for predictors or classifiers

Which text?
Which language?
Which model?
How to summarize?
Which questions to answer?
Converting unstructured texts into structured data (with NLP & ML)

Sample Text from a Mammography Report
There is a 1.8cm round mass with a circumscribed margin in the left breast in the anterior depth central to the nipple. Compared to previous films this mass is increase in size. There also is a 1.4cm oval mass with an obscured margin in the left breast in the anterior depth of the inferior region. Compared to previous films this mass is increased in size.

Output of an ideal information extraction system

Imaging Observation 1: Mass
- Size: 1.8cm
- Margin: Circumscribed
- Shape: Round
- Stability: Increasing
- Located In: left breast in the anterior depth central to the nipple.

Imaging Observation 2: Mass
- Size: 1.4cm
- Margin: Obscured
- Shape: Oval
- Stability: Increasing
- Located In: left the left breast in the anterior depth of the inferior region

Can we extract *missing* cancer stage data from clinical notes?

Stage information is *missing* from 10 to 50% of patient records in cancer registries.
NLP pipeline

Diagnosis Date  Surgery Date  60 days after surgery

Clinical Stage  Pathological Stage

All Narratives from Patient XX

Pre-processing

NLP Model

word2vec Trained on 1.4 million notes

Knowledge Base Creation

Clinical Notes

PathologyProgress reports

Pathology reports

Clinical Notes

Classification Model

Report - Key Term Vector Creation

Key Term Vector 1  Key Term Vector 2  ...

Report Vector = Average

NLP pipeline results


~30% Missing in EHR

Extracted
70% of missing pathological
30% of missing clinical stages from clinical text
Other examples: knowledge discovery from EHRs

Detecting the Timeline of Recurrence

Phenotyping Severity of PCOs

Detecting potential ambiguities

Examples

Metastatic recurrence of breast cancer

Urinary incontinence after chemo

Discrepancies between the laterality of an image and the reported impression.

NLP and ML
unlock the power of EHRs at scale

This involves making numerous careful decisions, rather than simply feeding all available data into a model and blindly accepting its output.
AI Governance

to ensure that these technologies are developed and deployed in a safe and responsible way
AI best practices in healthcare

As the number of models increases, it is becoming increasingly important to ensure that these models are fair, unbiased, & generalizable.

AI Governance in healthcare

Lifecycle and Key Dimensions of an AI System

A socio-technical team
Thank you!

Contact

selenb@stanford.edu
selenbw.github.io

I’m hiring!
Looking for interns, grad students and postdocs
ML-enabled data extraction, faster

Katherine Tan
Flatiron Health

@statistikat
wlktan
Katherine Tan is an employee of Flatiron Health, an independent subsidiary of Roche Group.

She holds stock ownership in Roche.
Cohort selection, identifying samples with greater specificity with respect to inclusion criteria

Identification of independent predictors and covariates of health outcomes

And much more…
Several data elements critical for outcomes research are stored in unstructured data sources.

Abstracting this information is a **costly** and **resource intensive** task.

---

**FOLLOW UP VISIT**

- **Reason for Visit / Chief Complaint**
  - Biopsy results.

- **Primary Diagnosis**
  - Secondary malignant neoplasm of other parts of nervous system (198.4/C79.49).

- **History of Present Illness**
  1. Metastatic adenocarcinoma of the lung identified in August. The patient reports that he developed back pain about 1 year prior to presentation. Pain waxed and waned over time, and he was treated with narcotic pain medicine. He presented to the emergency department for further evaluation of a pain exacerbation in June. He underwent plain films, which demonstrated degenerative changes and an osteoporotic wedge deformity at T9. This was not thought to be the source of his pain, since patient was reporting diffuse pain. He return to the emergency department in August with similar complaints, and underwent a chest x-ray and CT of the abdomen and pelvis. The CT scan demonstrated numerous bony lesions involving the lumbar spine and a probable pathologic fracture of the right iliac wing. Alkaline phosphatase was also elevated. PSA was normal at 2.1.
  2. History of peptic ulcer disease.
  3. Arthritis.

- **Interval History**
  - Returns to clinic today to review the results from his lymph node biopsy. He reports that he continues to have difficulty with back pain. It has improved some since his last visit with the assistance of palliative care; however, the back pain is still having a significant impact on his quality of life. He reports the pain is still 6-7 on a 10 point scale. He also continues to have some fatigue, but he reports that his activity level has improved some as his pain is improved. He reports generalized weakness, occasional cough, abdominal pain, poor appetite, and some intermittent numbness and tingling in his lower extremities. Remainder of his review of systems is documented below and was
Tradeoffs building EHR data solutions using traditional approach of manual chart review

**Scale and Size**
Limited cohort size, particularly for nuanced cohorts

**Recency and Speed**
Lag between research question and availability of temporally recent data

**Clinical Depth**
Relevant clinical details that are difficult to abstract at scale are missing
Deep learning models mimic human abstraction

- **Scalable, automated** extraction of clinical concepts **explicitly documented** in the patient chart

- **It is NOT**: prediction or inference of a clinical value based on other patient characteristics; generative modeling (ie, different from chat GPT)

---

<table>
<thead>
<tr>
<th>Deep Learning Model Example</th>
<th>Language in Source EHR as Illustrative Snippet (Model Input)</th>
<th>Extracted Variables (Model Outputs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomarkers</td>
<td>“Mr. Smith received NGS test results on 2/20/2021 for EGFR, ALK, and ROS1 and was found to have an ALK rearrangement.”</td>
<td>BiomarkerName, BiomarkerStatus, ResultReturnedDate</td>
</tr>
<tr>
<td>PD-L1</td>
<td>“Mr. Smith was diagnosed with adenocarcinoma of the lung, PD-L1 &lt;1% on 2/20/2021.”</td>
<td>BiomarkerName, PercentStaining, ResultReturnedDate</td>
</tr>
<tr>
<td>KRAS</td>
<td>“Mr. Smith tested positive for a KRAS G12C mutation on 1/15/2019.”</td>
<td>MutationG12CDetail</td>
</tr>
</tbody>
</table>

Adamson B, et al. Methods for machine learning extraction of RWD variables from electronic health records. medRxiv 2023.03.02.23286522
### Adding KRAS G12C mutation details to a lung cancer dataset using ML

In addition to **biomarker testing status**, knowing **mutation details** became important for researchers as SOC evolved and targeted therapies are approved.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Biomarker</th>
<th>Result date</th>
<th>G12C mutation?</th>
<th>Sample type</th>
<th>Tissue Collection site</th>
<th>Test type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>KRAS</td>
<td>2015-03-19</td>
<td>?</td>
<td>Tissue</td>
<td>Primary</td>
<td>NGS</td>
</tr>
<tr>
<td>B</td>
<td>KRAS</td>
<td>2017-01-23</td>
<td>?</td>
<td>Tissue</td>
<td>Metastatic</td>
<td>NGS</td>
</tr>
</tbody>
</table>
Name: John Smith
marker value result egfr (mutation) wild-type no mutation [**kras**] [**gl2c**] mutation **positive**

**EXAMPLE 1:**
**KRAS G12C**

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) [**BRAF**] [**V600E**] **positive**.

**EXAMPLE 2:**
**Other point mutation**

Table 1: summary of genes, values and results
Generalized Biomarker Models

EXAMPLE 1: KRAS G12C

Name: John Smith
marker value result egfr (mutation) wild-type no mutation [**kras**] [g12c] mutation positive ' table 1: summary of genes, values and results

EXAMPLE 2: Other point mutation

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) [**BRAF**] [v600e] positive. ' table 1: summary of genes, values and results

EXAMPLE 3: Generalized Biomarker Models

Name: Jane Doe
marker value result rrmi 0.37 low expression pdl1 high expression (60%) [**BIOMARKER**] [**MUTATION**] positive. ' table 1: summary of genes, values and results

EXAMPLE 4: Other point mutation

Name: John Smith
marker value result egfr (mutation) wild-type no mutation [**BIOMARKER**] [**MUTATION**] mutation positive ' table 1: summary of genes, values and results
ML-extraction enabled velocity of obtaining insights at scale and depth

**ABSTRACTION**

- 13,000 tasks
  - (3,700 hours)

**ML-EXTRACTION**

- 1,400 tasks
  - (10% of the abstraction load)
ML-extracted data can generate similar results and conclusions as abstracted data

Results from replication of natural history study of biomarker associated survival


Thank you

Additional Collaborators: Erin Fidyk, Blythe Adamson, Chaya Wurman, Melissa Estevez, Sheila Nemeth, Catherine Au-Yeung (Design)

Katherine Tan
Senior Quantitative Scientist
Machine Learning & Data Capabilities
Flatiron Health
@statistikat
wlktan
Why Human-Algorithm Collaborations will Transform Care Delivery

Ravi B. Parikh, MD, MPP
Ravi.parikh@pennmedicine.upenn.edu
@ravi_b_parikh
Disclosures

- **Research Funding:** NIH, Department of Defense, Veterans Health Administration, National Palliative Care Research Center, Humana, Prostate Cancer Foundation, NCCN Foundation, Conquer Cancer Foundation, Emerson Collective
- **Consultant:** Thyme Care, Onc.ai, Biofourmis, Humana, Cancer Study Group, GNS Healthcare, Nanology
- **Columnist:** Medscape, Flatiron
- **Leadership:** Coalition to Transform Advanced Care, National Quality Forum
My first experience with AI

Readmission Risk

| JENSEN, SERGIO | PT VISIT | 60min |
| MOSLEY, MALAYA | ADULT SICK VISIT | 15min |
| BLACKWELL, CESAR | PT VISIT | 60min |
| CHRISTENSEN, KYRKE | CHILD SICK VISIT | 30min |
| HODGES, ELLIOTT | PT VISIT | 60min |
| REESE, CALEB | CONSULTATION | 30min |
| KNIGHT, BLAKE | NEW PT VISIT | 60min |
| WOODWARD, ROCCO | CONSULTATION | 30min |
| WALKER, MARA | NEW PT VISIT | 60min |
| PETERSEN, EVIE | CHILD SICK VISIT | 30min |
| CROSS, KEATON | TELEMEDICINE | 30min |
| MCKEE, ANABELLA | ADULT SICK VISIT | 15min |
| STOUT, LIBBY | PT VISIT | 60min |
| JORDAN, LONDON | PT RE-EVAL | 20min |
| HATFIELD, ESTEBAN | PT VISIT | 60min |
| CHANAY, LINDSAY | CONSULTATION | 30min |
| CAMERON, CORBIN | PT VISIT | 60min |

<table>
<thead>
<tr>
<th>Sign</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROACH, TRISTIN</td>
<td>Fibrinogen, INR, PT, TTT</td>
</tr>
<tr>
<td>ROACH, TRISTIN</td>
<td>Lipitor 80 mg</td>
</tr>
<tr>
<td>LEON, ERIN</td>
<td>Geriatric Wellness Visit</td>
</tr>
<tr>
<td>BECK, ALVIA</td>
<td>Zocor 20 mg</td>
</tr>
<tr>
<td>NORTON, BETHANY</td>
<td>Norvasc 10 mg</td>
</tr>
<tr>
<td>MONTGOMERY, BLAINE</td>
<td>GlucoPhage 850 mg</td>
</tr>
<tr>
<td>KLECK, MICHAEL</td>
<td>Office Visit - Abbreviated</td>
</tr>
<tr>
<td>MCARDLE, HELEN</td>
<td>Office Visit - Mobile</td>
</tr>
<tr>
<td>BERN, MARC</td>
<td>Office Visit - Itemized Conditions</td>
</tr>
<tr>
<td>ANDERSON, JIM</td>
<td>Advanced Directives</td>
</tr>
<tr>
<td>BECKER, JOSEPH</td>
<td>Office Visit1</td>
</tr>
<tr>
<td>HANSEN, GEORGE</td>
<td>Office Visit1</td>
</tr>
<tr>
<td>FALK, MICHAEL A</td>
<td>Urine Albumin/Creatinine, Urine C &amp; S</td>
</tr>
<tr>
<td>FERNANDEZ, MEGAN</td>
<td>Urine Albumin/Creatinine, Urine C &amp; S</td>
</tr>
<tr>
<td>DEAN, BRIAN</td>
<td>25(OH)D, ANA, B12, C &amp; S, CRP, ESR or Sedrat...</td>
</tr>
<tr>
<td>CAMPBELL, LISA C</td>
<td>Blood Urea Nitrogen, Calcium, Carbon Dioxide, Ch...</td>
</tr>
<tr>
<td>BECKER, JOSEPH</td>
<td>#186</td>
</tr>
</tbody>
</table>
We pretend like all AI is autonomous…

<table>
<thead>
<tr>
<th>Assistive AI algorithms</th>
<th>Autonomous AI algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td><strong>Level 2</strong></td>
</tr>
<tr>
<td>Event monitoring</td>
<td>Event monitoring</td>
</tr>
<tr>
<td>AI</td>
<td>AI</td>
</tr>
<tr>
<td>Response execution</td>
<td>Response execution</td>
</tr>
<tr>
<td>Clinician</td>
<td>Clinician and AI</td>
</tr>
</tbody>
</table>

**Deep neural network approach**

- **Input**: ECGs
- **Layer 1**: Convolutional
- **Layer 2**: Fully connected
- **Output**: Low EF
  - Type of heart rhythm:
    - Atrial fibrillation
    - Atrioventricular block
    - Supraventricular tachycardia

@ravi_b_parikh

Michole and Rodriguez, Nat Med, 2019
...when instead most current AI is assistive

<table>
<thead>
<tr>
<th>Assistive AI algorithms</th>
<th>Autonomous AI algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td><strong>Level 3</strong></td>
</tr>
<tr>
<td>Data presentation</td>
<td>Conditional automation</td>
</tr>
<tr>
<td>Event monitoring</td>
<td>AI</td>
</tr>
<tr>
<td>Clinician</td>
<td>AI</td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td><strong>Level 4</strong></td>
</tr>
<tr>
<td>Clinical decision-support</td>
<td>High automation</td>
</tr>
<tr>
<td>Response execution</td>
<td>AI</td>
</tr>
<tr>
<td>Clinician</td>
<td>AI</td>
</tr>
</tbody>
</table>

@ravi_b_parikh
Human-algorithm collaborations require more than data science

Aim 1: Human inputs
- High variability
- Cognitive biases

Aim 2: Human inputs
- High variability
- Cognitive biases

Aim 3: Contextual inputs
- Alarm fatigue
- Inadequate resources to act on a prediction

Human-Machine Collaboration

Accuracy? Trust?
Impact on decision-making?

Predictions
Decision-Making

Machine inputs
- Inadequate access to all information sources
- Poor explainability
A use case: Serious Illness Communication

- Early communication is key to reducing unwarranted end-of-life care in oncology
- Identifying appropriate patients is key, but oncologists do badly at prognosis
Why do existing solutions fail in real-world practice?

- Identifying the right patients is hard
- Changing behavior is hard

Machine Learning

Behavioral Economics

Christakis and Lamont, BMJ, 2003; Bestvina and Polite, JOP, 2017
Algorithm development and validation

### Variables

#### Demographics
- Age, Gender

#### Comorbidities
- 33 Elixhauser comorbidities
  - Total count
  - Recent*

#### Cancer-specific
- Stage, tumor markers
  - Total count
  - First/last value
  - Min/Max
  - Proportion ordered STAT

#### Laboratories
- CMP, CBC, LDH

#### Recent utilization
- Outpatient visit number
Incorporating behavioral economics with machine learning

**Use Case:** Predicting mortality to prompt more serious illness communication

Next week’s high-risk patients for Serious Illness Conversations

Thursday, July 18, 2019 at 8:05 AM
Show Details

Dear:

The ACC is working to help oncologists have earlier Serious Illness Conversations with patients. In the past four weeks, you have documented 2 conversations.

18 oncology clinicians have documented more conversations than you during that time.

We have identified patients scheduled to see you next week who may benefit from a Serious Illness Conversation. [Click here](#) to view your list (you must be connected to the UPHS network).

Sincerely,

Peer Comparison

Performance Report

ML high-risk list

Pre-commitment

Default Text Message

Parikh et al, JAMA Onc, 2022

@ravi_b_parikh
Conversational Connect Impact

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemo last 14 days*</td>
<td>10.4%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Hospice before death</td>
<td>59.6%</td>
<td>60.6%</td>
</tr>
<tr>
<td>ICU last 30 days</td>
<td>16.9%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Spending last 6 months of life*</td>
<td>$65,971.09</td>
<td>$78,256.24</td>
</tr>
</tbody>
</table>
Humans don’t respond to AI in the same way

Lower-volume oncologists are more likely to respond to a machine learning nudge

Phenotyping clinicians can help refine AI interventions

Li et al, PLOS One, 2022
Human-centered AI can mitigate disparities

<table>
<thead>
<tr>
<th></th>
<th>Pre-intervention, %</th>
<th>Post-intervention, %</th>
<th>Absolute Percentage-point Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-Hispanic White</td>
<td>3.9 (58/1494)</td>
<td>14.2 (201/1417)</td>
<td>10.3</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>3.6 (17/467)</td>
<td>16.9 (69/408)</td>
<td>13.3</td>
</tr>
<tr>
<td>Other*</td>
<td>1.2 (2/164)</td>
<td>19.5 (34/408)</td>
<td>18.3</td>
</tr>
</tbody>
</table>
Human-in-the-loop models are a promising strategy
Lessons Learned

Machine learning predictions can improve care when

- Clinicians’ perspectives are solicited prior to algorithm development
- The algorithm is “vetted” prior to implementation
- They are well-integrated into clinical workflows
- They are paired with behavioral nudges rather than simply displayed on a computer screen