Computational Health for Next Generation Healthcare

Jianying Hu

Global Science Leader, AI for Healthcare
Program Director, Center for Computational Health
IBM Fellow
Precision Medicine

Prevention and treatment strategies that take individual variability into account

- Dr. Francis Collins, NIH Director
Computational Health: From Data to Impact

Pattern Extraction
- Preprocessing
- Feature Engineering
- Feature Selection

Insights Generation
- Precision Cohort
- Complex Models
- Longitudinal Modeling
- Interpretable Models
- Causal Inference

Personalized Intervention
- Reasoning and Decision Support
- Behavioral Profiling and Intervention
- Patient Centric Measures and Assessments

Knowledge Sources
Patient Similarity Analytics for Precision Cohort

Goal

– Identify patients who are similar to a given patient of interest in a clinically meaningful way
– Identify a measure of clinical similarity between patients

Approach

– Supervised metric learning

Challenges Addressed

– Patient similarity is context dependent
– Feature dimensionality can be very large

Healthcare Predictive Modeling Framework

- ML based feature preprocessing to handle incompleteness and sparsity
- Pattern mining to derive higher level phenotype representations
  - Advanced feature selection algorithms to identify salient signals from high dimensional data
  - Big data platform to support rapid exploration of large model space

Disease areas applied to: CHF, diabetes, hypertension, schizophrenia ...
Heart Failure Predictive Modeling

- 4644 case patients, 45,981 control patients
- Over 20,000 features of different types (diagnoses, demographics, Framingham symptoms, lab results, medications, vitals)
- Novel feature selection algorithm enabling integration of knowledge driven and data driven risk factors

$2$ Million Awarded to Sutter Health, IBM and Geisinger Health System to Study Heart Failure Prediction
Three-year collaboration will develop groundbreaking, big data analytic methods to improve care and reduce costs for treatment of heart disease

Personalized Predictive Models - T2D Onset Prediction Example

- Diabetes patient population is heterogeneous
- Traditional predictive modeling approaches only provide “universal” risk factor identification and ranking
- Personalized predictive modeling approach provides patient specific risk factors and ranking

Published in AMIA Joint Summit 2015
Comprehensive Risk Assessment – Multi-Task Sparse Learning

Goal

- Simultaneously predict multiple risks
- Explore and exploit risk associations
- Identify common and unique risk factors

Use Cases: elder care risk assessment, diabetes complications

Published in KDD 2015, AAAI 2018
Disease Progression Modeling

• **GOAL** → Provide comprehensive view and deeper understanding of a disease in terms of characteristics of underlying disease stages, areas of manifestation and progression pathways

• **METHOD** → Multi-layer probabilistic modeling framework to incorporate data from diverse sources, account for incomplete data, extract salient phenotypes, infer (potentially hidden) states and transition probabilities.

• Initial work on COPD; Work ongoing on enhanced methodologies & application to other conditions, including Huntington’s (CHDI), T1D (JDRF), PD (MJFF).

Published in KDD 2014, AMIA 2017, AMIA 2018, JAMIA Open 2019
Huntington’s Disease Progression Modeling

Challenges in Understanding HD Progression

- Disease manifestation along multiple dimensions with complex patterns
- Heterogeneous progression pathways
- No clear definitions of disease states

Probabilistic Disease Progression Modeling

- Incorporate heterogeneous features coming from multiple studies and assessments covering multiple aspects of HD
- Provide comprehensive view of disease states and the progression of HD through a multi-layer probabilistic disease progression model
- Better understanding of disease sub-types
- Identify factors that are associated with disease progression patterns

Better Management and Care for Patients

- Improved understanding of disease progression: population/patient
- Insights into HD clinical assessments and sensitivities

More Effective Drug Discovery

- Objective baseline
- Identification of the most appropriate cohort – trial enrichment
- Optimization of trial design – trial simulator
- Knowledge regarding landmark non-motor events and their temporal relationship with clinical diagnosis
Integrated Huntington’s Disease Progression Model

- Trained on data integrated from four prospective observational studies of HD (~16k case, 3k control)
- Discovered 9 disease states, over span of ~4 decades (prodromal, transition, manifest)
- Highlights (compared to prior-art HD progression indices)
  - Capturing multi-faceted manifestation throughout disease progression
  - Finer characterization, particularly of early states
  - Characterization of complex patterns of progression in transition (critical) states
  - Individual patient: more nuanced view of state of progression

Published in JAMIA Open 2019
Causal Inference for Time Varying Treatment Strategies

- **Goal**: To provide causal estimate of effect of time varying treatments using observational data.
- **Method**: G-Computation + hierarchical Bayesian models (post-processing).
- **Challenges addressed**:
  - Time varying treatment decisions
  - Outcomes recorded at irregular intervals/varying treatment durations
  - Multiple related drugs and multiple related outcomes

The G-formula:

\[
E[Y_t(g)] = \sum_{[I_t]} E[Y_t | \bar{I}_t = \bar{I}_t, \bar{A}_t = g(\bar{I}_t)] \prod_{m=1:t} p(L_m = l_m | \bar{I}_{m-1} = \bar{I}_{m-1}, \bar{A}_{m-1} = g(\bar{I}_{m-1}))
\]

- Model for outcome given the past
- Models for confounders given the past

Published in AMIA Informatics Summits 2019
Deep Learning Models for Predicting Epileptic Seizures

- **GOAL**: Build algorithms that allow a patient to manage their condition, alerting them to impending seizures.
- **METHOD**: Real-Time Classifications Using AI – Neural Network Architecture to recognize patient-specific patterns emerging before a seizure.
- **DATA**: Long-term intracranial EEG recordings from 15 patients provided by St. Vincent’s Hospital labelled by neurologists.

Train algorithm to distinguish between pre-seizure & normal brain signal.
Enable prediction up to 5 hours away from seizure.

System allows for real-time alarms & tuning based on patient’s needs:
- Network classifies each sample every **30 seconds**.
- Without post-processing, this would result in **one prediction every 30 seconds**.
- Exploiting the sequential structure of the data, temporal averaging should improve classification accuracy.

Published in EBioMedicine 2018
Predictive Modeling with Deep Learning from EMR

- Temporal patterns and interactions are important features in predictive modeling in healthcare.
- Prior methods do not sufficiently address the challenge of extracting such features from longitudinal patient record matrices (EHR).
- We developed an end-to-end deep learning framework tailored to longitudinal health care data to:
  - Learn the temporal patterns
  - Exploit them for risk prediction

Disease Onset Prediction Results (AUC)

<table>
<thead>
<tr>
<th></th>
<th>CHF</th>
<th>COPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.7156</td>
<td>0.6624</td>
</tr>
<tr>
<td>Deep Learning</td>
<td>0.7675</td>
<td>0.7388</td>
</tr>
</tbody>
</table>

Published in SDM 2016
Home Monitoring of Parkinson’s Disease Patients

Connected sensors, mobile devices, & machine learning provide real-time, around-the-clock disease symptom information to clinicians and researchers.

Working to transform Parkinson’s Disease care

1 million American currently have Parkinson’s
4% are under 50 years of age
$25 billion in annual direct and indirect costs of care, social payments and lost income
Home Monitoring: Parkinson’s Disease Applications

**GOAL** ➔ Movements captured passively & continuously in an IoT-enabled environment to improve outcomes for Parkinson’s Disease patients

**APPROACH** ➔ Complex human movements decomposed into movement primitives to classify motor assessment, with real-time correlation

Precision Stress Management

- **GOALS** - Passively identify stressful events in the lab and field using physiological measurements collected from consumer grade device(s)
- Use reliable stress indicators to guide timing and nature of behavioral interventions

**KEY FINDINGS**

**In-Lab:**
Heart rate variability measured using PPG sensor provided the most correlated signal to perceived stress, compared to other physiological signals (e.g., EEG, HR, EDA)

**In-Field:**
IBM Stress Model developed using PPG data was > 85% accurate in identifying stress events

Published in AMIA 2017, MedInfo 2017
AI for Healthcare: Accelerating the Journey

Knowledge

Learning from all types of data, together
Probabilistic temporal models
Distributed learning

Learning from small data
Enhanced causal inference
Interpretable models

Comprehensive models for continuous understanding, prediction and reasoning

- Learn characteristics and drivers of diseases and their progressions
- Predict health encounters and outcomes
- Reason about personalized treatments and interventions

All Patient Data
- clinical
- genomics
- physiological
- imaging
- lifestyle
- social
- behavioral

Novel AI Algorithms for Learning and Reasoning

Proactive Care – Deeper Clinical Insight
Better Outcomes