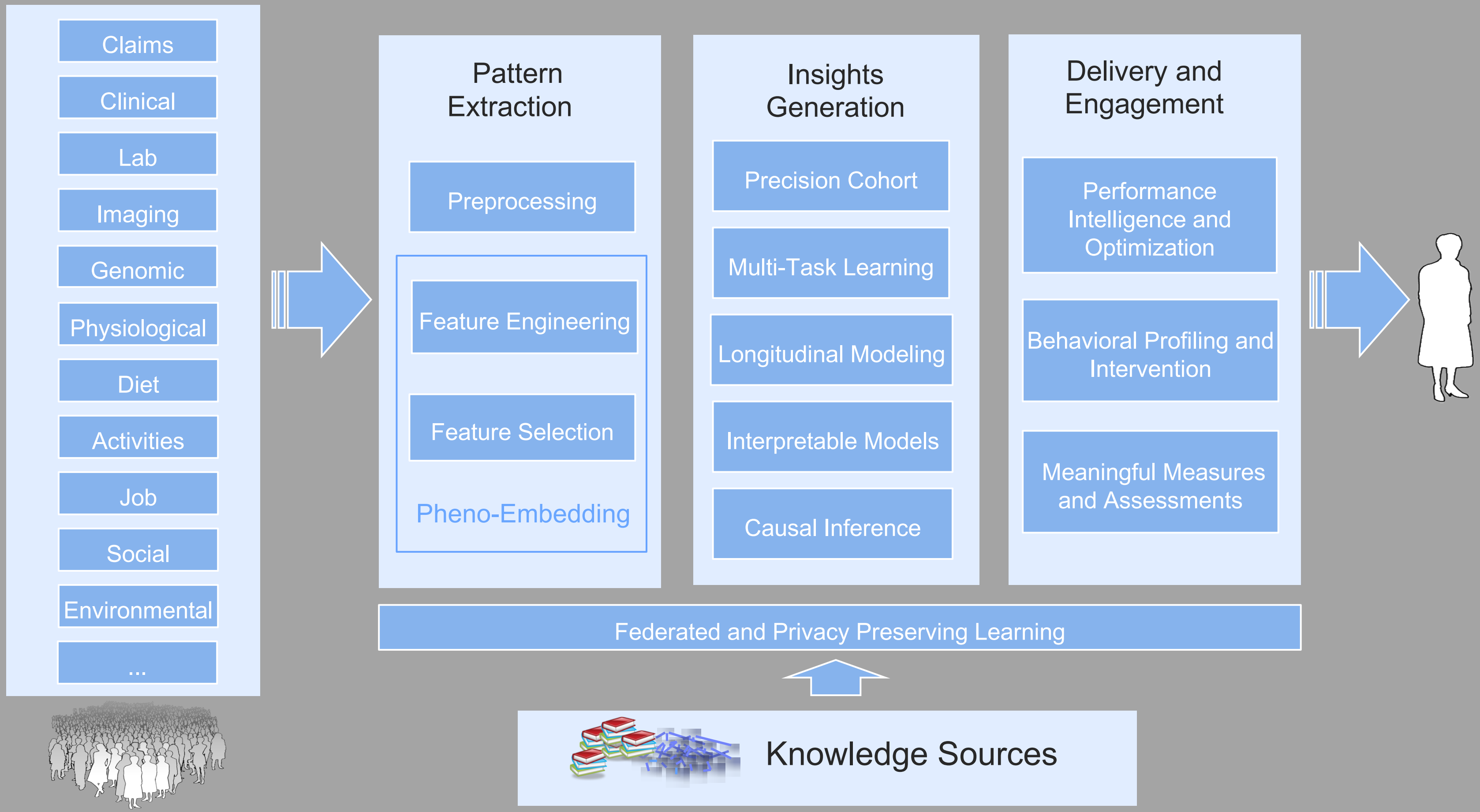
Computational Methods for Next Generation Healthcare

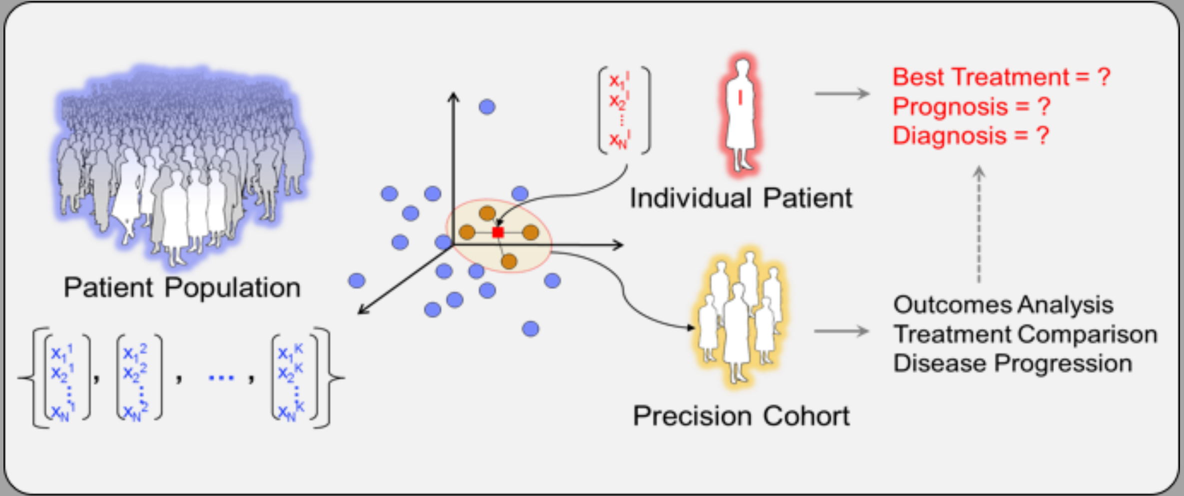
Jianying Hu

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

Computational Health: From Data to Impact



Patient Similarity Analytics for Precision Cohort



<u>Goal</u>

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

<u>Approach</u>

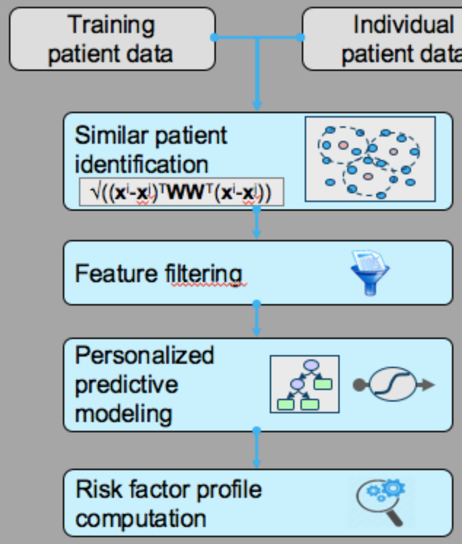
- Supervised metric learning

<u>Challenges Addressed</u>

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

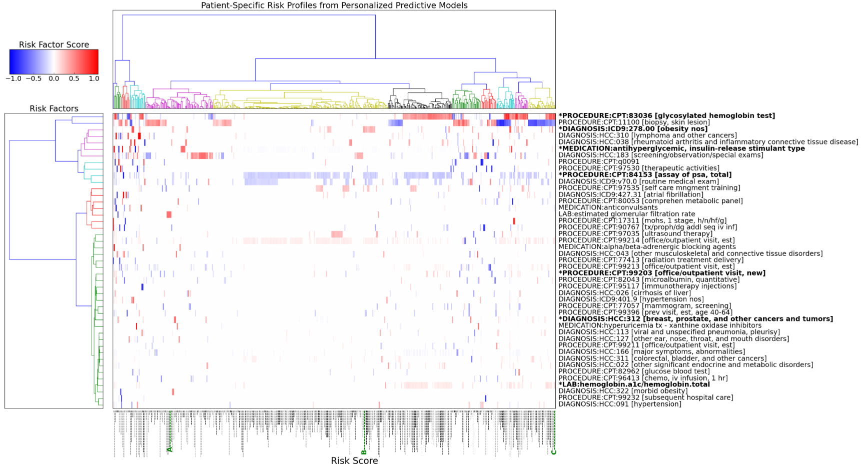
Published in: AMIA 2010, ICDM2010, SDM2011, ICPR2012, AMIA TBI 2014, AMIA CRI 2015

Personalized Predictive Models - T2D Onset Prediction Example



Insights for personalized intervention planning

- ✓ 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
- ✓ Clusters of risk factors, and patient risk profiles



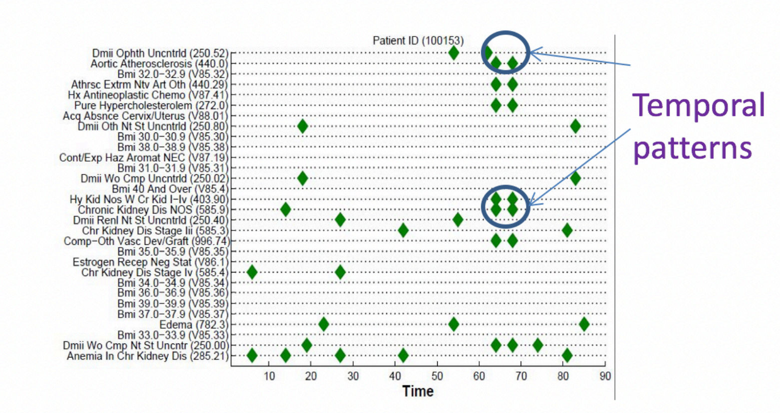
AMIA Joint Summit 2015

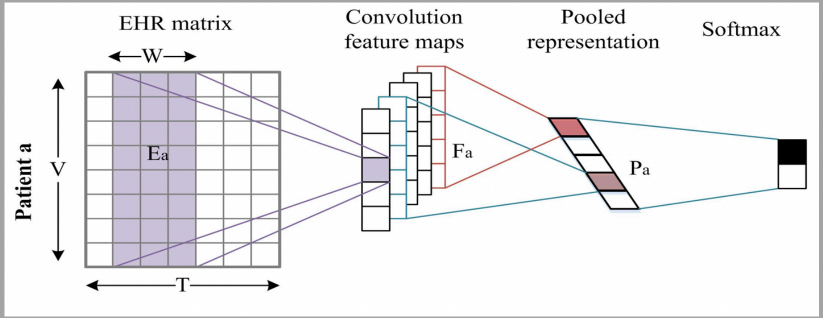
Temporal Pattern Extraction with Deep Learning from EMR

- 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 pattern and exploit them for risk prediction

Disease Onset Prediction Results (AUC)

	CHF	COPD
Baseline (Random Forest)	0.7156	0.6624
Deep Learning (Slow-Fusion Convolutional Neural Network)	0.7675	0.7388

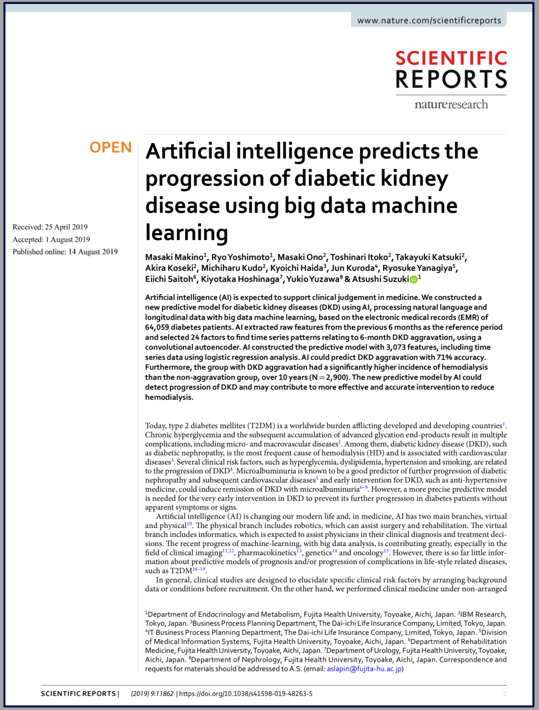




SDM 2016

Diabetic Kidney Disease Prediction

FUJITA HEALTH UNIVERSITY



www.nature.com/scientificreports

SCIENTIFIC REPORTS

natureresearch

OPEN

Received: 25 April 2019
Accepted: 1 August 2019
Published online: 14 August 2019

Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning

Masaki Makino[1], Ryo Yoshimoto[1], Masaki Ono[2], Toshinari Itoko[2], Takayuki Katsuki[2], Akira Koseki[2], Michiharu Kudo[2], Kyoichi Haida[3], Jun Kuroda[4], Ryosuke Yanagiya[5], Eiichi Saitoh[6], Kiyotaka Hoshinaga[7], Yukio Yuzawa[8] & Atsushi Suzuki[1]

Artificial intelligence (AI) is expected to support clinical judgement in medicine. We constructed a new predictive model for diabetic kidney diseases (DKD) using AI, processing natural language and longitudinal data with big data machine learning, based on the electronic medical records (EMR) of 64,059 diabetes patients. AI extracted raw features from the previous 6 months as the reference period and selected 24 factors to find time series patterns relating to 6-month DKD aggravation, using a convolutional autoencoder. AI constructed the predictive model with 3,073 features, including time series data using logistic regression analysis. AI could predict DKD aggravation with 71% accuracy. Furthermore, the group with DKD aggravation had a significantly higher incidence of hemodialysis than the non-aggravation group, over 10 years (N = 2,900). The new predictive model by AI could detect progression of DKD and may contribute to more effective and accurate intervention to reduce hemodialysis.

Today, type 2 diabetes mellites (T2DM) is a worldwide burden afflicting developed and developing countries[1]. Chronic hyperglycemia and the subsequent accumulation of advanced glycation end-products result in multiple complications, including micro- and macrovascular diseases[2]. Among them, diabetic kidney disease (DKD), such as diabetic nephropathy, is the most frequent cause of hemodialysis (HD) and is associated with cardiovascular diseases[3]. Several clinical risk factors, such as hyperglycemia, dyslipidemia, hypertension and smoking, are related to the progression of DKD[4]. Microalbuminuria is known to be a good predictor of further progression of diabetic nephropathy and subsequent cardiovascular diseases[5] and early intervention for DKD, such as anti-hypertensive medicine, could induce remission of DKD with microalbuminuria[6–9]. However, a more precise predictive model is needed for the very early intervention in DKD to prevent its further progression in diabetes patients without apparent symptoms or signs.

Artificial intelligence (AI) is changing our modern life and, in medicine, AI has two main branches, virtual and physical[10]. The physical branch includes robotics, which can assist surgery and rehabilitation. The virtual branch includes informatics, which is expected to assist physicians in their clinical diagnosis and treatment decisions. The recent progress of machine-learning, with big data analysis, is contributing greatly, especially in the field of clinical imaging[11,12], pharmacokinetics[13], genetics[14] and oncology[15]. However, there is so far little information about predictive models of prognosis and/or progression of complications in life-style related diseases, such as T2DM[16–19].

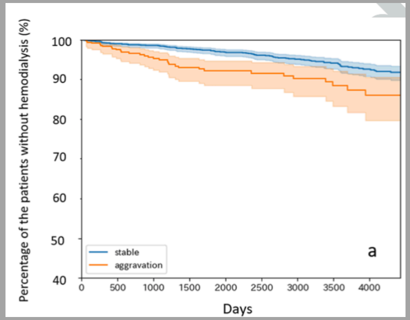
In general, clinical studies are designed to elucidate specific clinical risk factors by arranging background data or conditions before recruitment. On the other hand, we performed clinical medicine under non-arranged

[1]Department of Endocrinology and Metabolism, Fujita Health University, Toyoake, Aichi, Japan. [2]IBM Research, Tokyo, Japan. [3]Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Tokyo, Japan. [4]IT Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Tokyo, Japan. [5]Division of Medical Information Systems, Fujita Health University, Toyoake, Aichi, Japan. [6]Department of Rehabilitation Medicine, Fujita Health University, Toyoake, Aichi, Japan. [7]Department of Urology, Fujita Health University, Toyoake, Aichi, Japan. [8]Department of Nephrology, Fujita Health University, Toyoake, Aichi, Japan. Correspondence and requests for materials should be addressed to A.S. (email: aslapin@fujita-hu.ac.jp)

- Deep Learning to extract temporal features
- Logistic regression for DKD prediction

Features	AUC	Accuracy
Profile	0.562	0.548
Profile + ICD10	0.562	0.557
Profile + ICD10 + YJCode	0.613	0.594
Profile + ICD10 + Blood Tests (latest)	0.644	0.606
Profile + ICD10 + YJCode +Blood Tests (latest and longitudinal)	0.656	0.610
Profile + ICD10 + YJCode +Blood Tests (latest and longitudinal) +Urinary Tests (latest and longitudinal)	0.729	0.691
Profile + ICD10 + YJCode +Blood Tests (latest and longitudinal) +Urinary Tests (latest and longitudinal) +Current Disease + Disease History	0.743	0.701

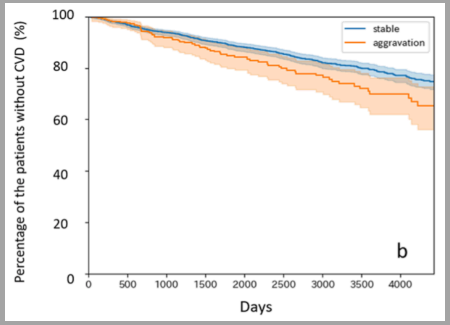
Impact of data coverage on performance



Survival analysis for hemodialysis

Key Findings

- ✓ Additional data categories improve prediction accuracy
- ✓ The aggravation of urinary protein observation is strongly affected by its variance over past 180 days
- ✓ DKD aggravation group had significantly higher incidence rates of Hemodialysis and Cardiovascular diseases

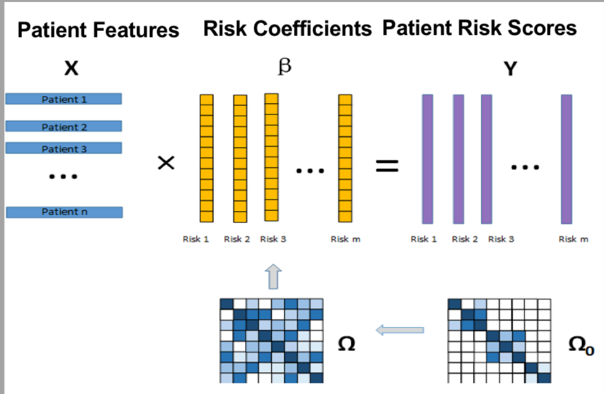


Survival analysis for CVD

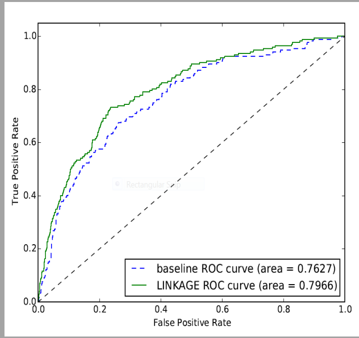
Comprehensive Risk Assessment – Multi-Task Sparse Learning

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

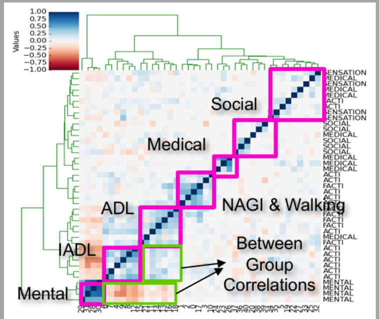
Use Cases: elder care risk assessment, diabetes complications



$$\min_{\beta,\Omega} \; l(Y,X,\beta) + \mathrm{tr}\left[\left(\frac{\lambda_1}{2}\beta^\top\beta + \frac{\lambda_2}{2}\Omega_0\right)\Omega^{-1}\right] + \frac{\lambda_3}{2}\log\det(\Omega)$$
$$+\,\gamma\sum_{i\neq j}|\Omega_{ij}|\,\|\beta_i - \mathrm{sign}(\Omega_{ij})\,\beta_j\|_1 + \gamma_2\|\beta\|_1\,,$$
$$s.t.\ \Omega \succeq 0.$$



Improved Prediction Accuracy



Identify Association Among Risks

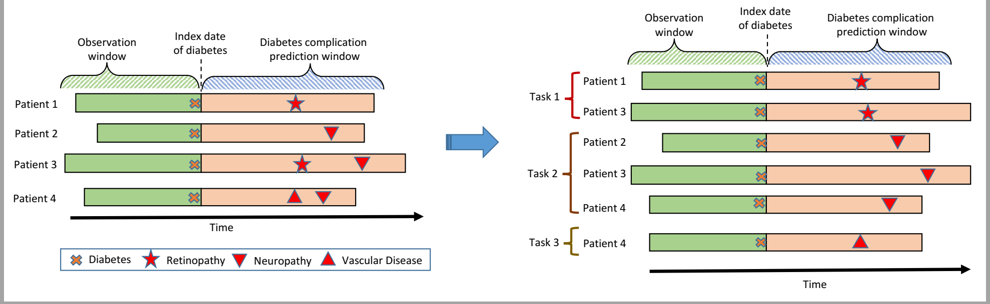
KDD 2015

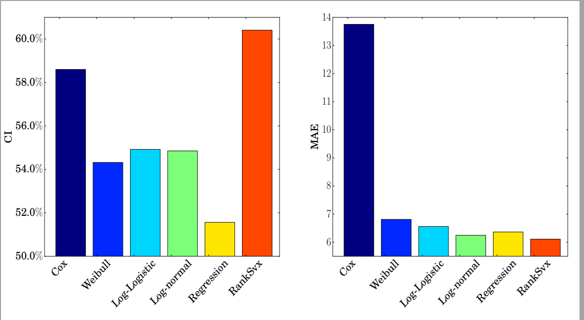
Multi-Task Learning for Diabetes Complications Prediction

Goal : Predict ***when*** a patient will develop complications after the initial T2DM diagnosis

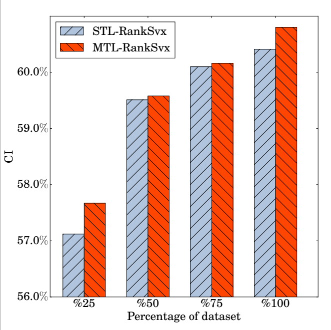
Approach : Multi-task Survival Analysis

- **RankSvx**: A novel data-driven time-to-event modeling method
  - Accurate prediction of event times, and
  - Ranking of the relative risks among patients
- **Multi-task RankSvx** to simultaneously model multiple complication risks
  - leverage association between different diabetes complications
- Applied to predictions of retinopathy, neuropathy, nephropathy, vascular diseases





RankSvx outperforms traditional survival models and regression model in CI and MAE

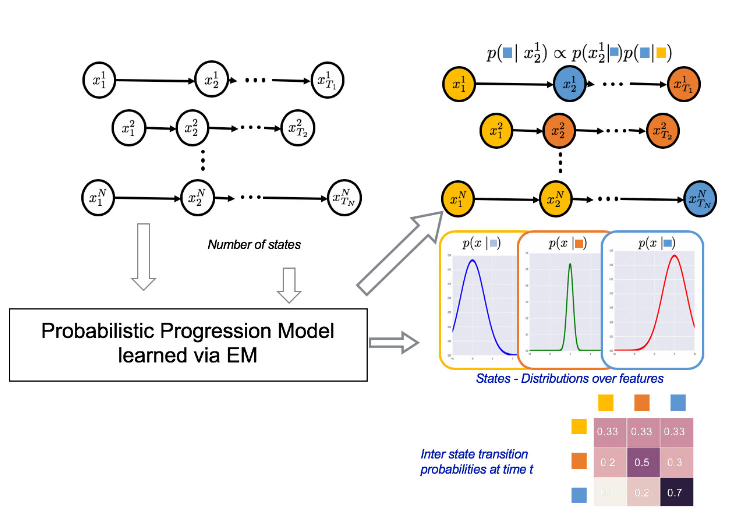


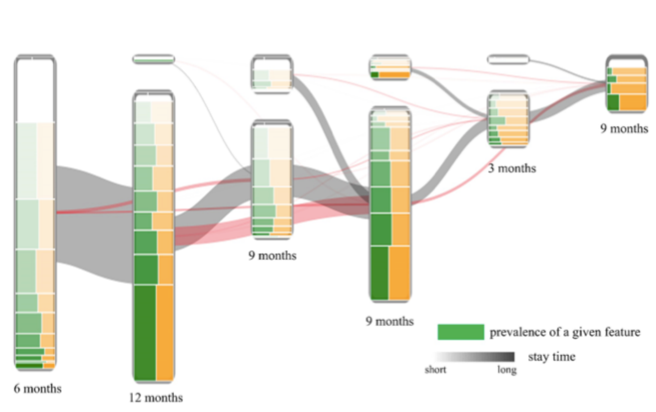
MTL-RankSvx outperforms STL-RankSvx

AAAI 2018, IEEE TKDE 2019

Disease Progression Modeling

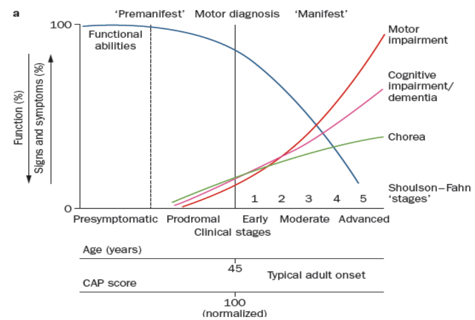
- GOAL → Provide comprehensive view and deeper understanding of of a disease in terms of characteristics of underlying disease stages, areas of manifestation and progression pathways
- METHOD → Multi-layer probabilistic modeling framework to incorporate data from diverse sources
- Initial work on COPD; 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



Natural History of HD (Ross et. at. 2014)

Challenges in Understanding HD Progression

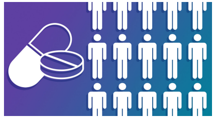
- Disease manifestation along multiple dimensions with complex patterns
- Heterogeneous progression pathways
- No clear definitions of disease states

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

Clinical Decision Support

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



Clinical Trial Design

- ✓ Objective baseline
- ✓ Cohort selection – trial enrichment
- ✓ Optimizing trial design – trial simulator
- ✓ Biomarker discovery

Integrated Huntington's Disease Progression Model

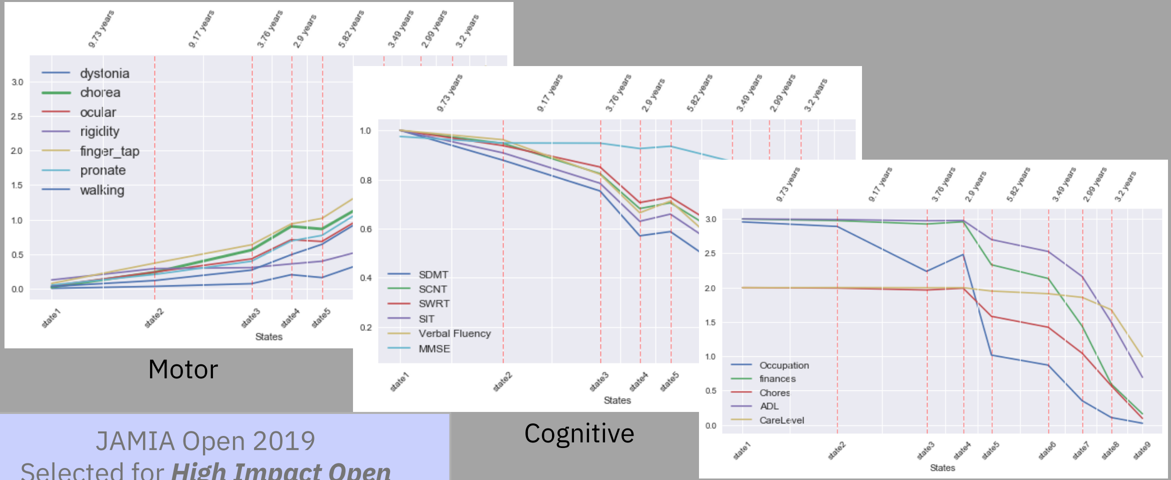
- Trained on data integrated from four prospective observational studies of HD (~16k case, 3k control)
- Discovered 9 disease states, over span of ~4 decades (prodromal, transition, manifest)
- Highlights (compared to prior-art HD progression indices)
  - Capturing multi-faceted manifestation throughout 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

Annual transition probabilities through successive phenotypes ranges from 5% - 30%

Annual

Stage	State	Prodromal 1	Prodromal 2	Transition 3	Transition 4	Transition 5	Manifest 6	Manifest 7	Manifest 8	Manifest 9
Prodromal	1	0.93	0.057	0.0021						
Prodromal	2		0.81	0.13	0.048	0.013	0.0059			
Transition	3			0.68	0.15	0.12	0.041	0.005		
Transition	4				0.72	0.1	0.15	0.022	0.0026	
Transition	5					0.63	0.3	0.056	0.0075	
Manifest	6						0.75	0.21	0.036	0.0051
Manifest	7							0.72	0.24	0.048
Manifest	8								0.73	0.27
Manifest	9									1

Population view of multiple aspects of disease progression of HD



Motor

Cognitive

Functional

Table 2. State sequence of an example patient

Visit date (years)	State from IHDPM	Shoulson and Fahn Stage
0	2	Premanifest
1.1	2	Premanifest
2.1	2	Premanifest
3.5	2	Premanifest
4.2	2	Premanifest
5.7	2	Premanifest
6.6	3	Premanifest
7.5	3	Premanifest
8.9	3	HD1
10.3	4	HD1
11.4	4	HD1
11.8	4	HD1
13.6	5	HD1
14.7	6	HD2

JAMIA Open 2019
Selected for ***High Impact Open Access Research*** Collection in 2020

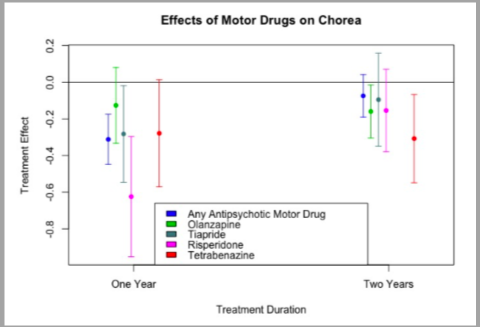
Causal Inference for Time Varying Treatment Strategies

IBM CHDI FOUNDATION

- 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:

$$E[Y_t(g)] = \sum_{\{\bar{l}_t\}} \underbrace{E[Y_t|\bar{L}_t = \bar{l}_t, \bar{A}_t = g(\bar{l}_t)]}_{\text{model for outcome given the past}} \prod_{m=1:t} \underbrace{p(L_m = l_m|\bar{L}_{m-1} = \bar{l}_{m-1}, \bar{A}_{\{m-1\}} = g(\bar{l}_{m-1}))}_{\text{models for confounders given the past}}$$



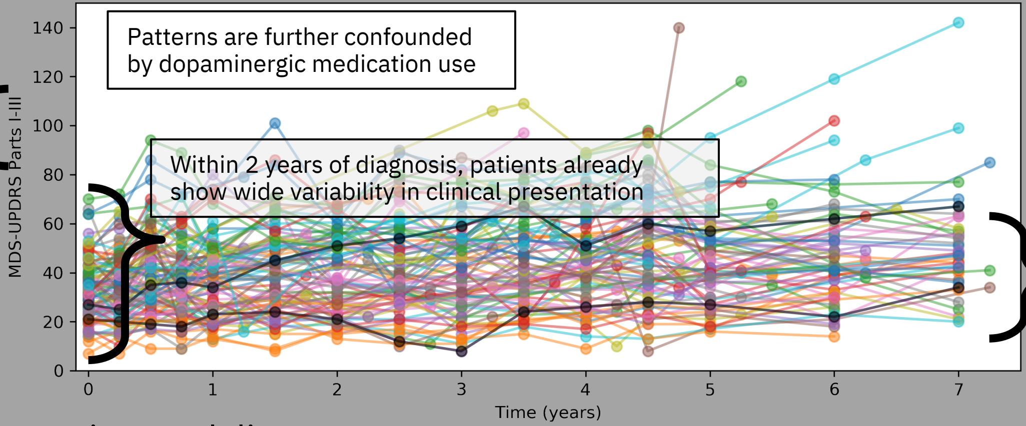
Modeling of Parkinson’s Disease Progression

IBM

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON’S RESEARCH

Challenges in Understanding PD Progression

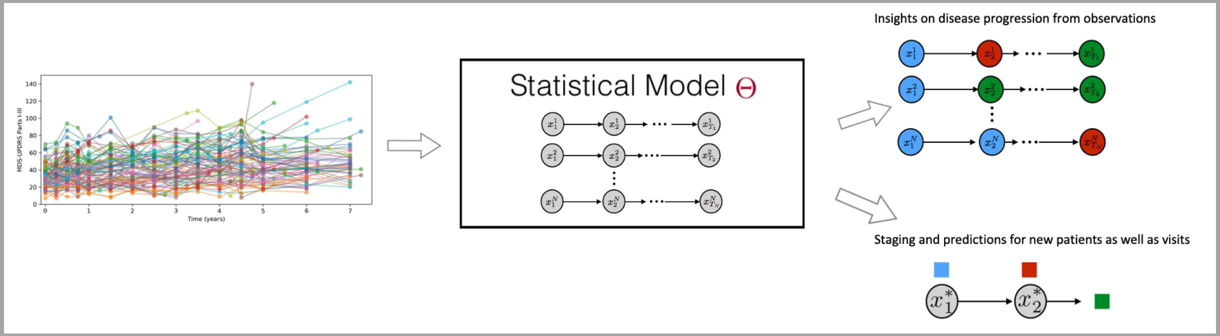
Summary scores mask symptom patterns



Inconsistent trajectory patterns across patients, even when controlling for early clinical presentation

Probabilistic Disease Progression Modeling

- Incorporate many clinical measures of PD
- Use control cases to subtract away non-PD effects (e.g. aging)
- Use input-outcome HMM approach to train personalized Medication Aware progression models



AAAI 2019, MLHC 2020

Improve Patient Engagement in Care Management

Engaging patients in interventions that are most effective for patients like them

Objectives

- Help care managers prioritize patients who will be more receptive to care management interventions.
- Help care managers set behavior goals/interventions based on intervention effectiveness estimates.

Data

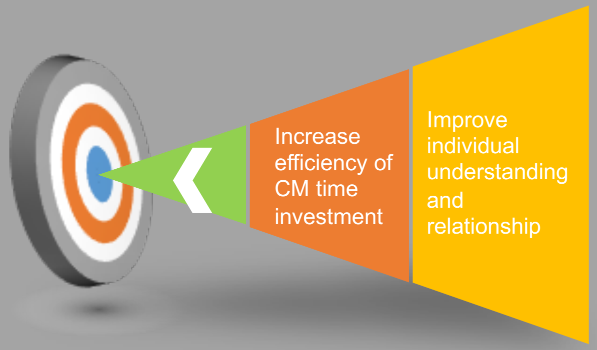
Care management records (structured + unstructured)

Method

Patient similarity, Causal inference,

Key Finding

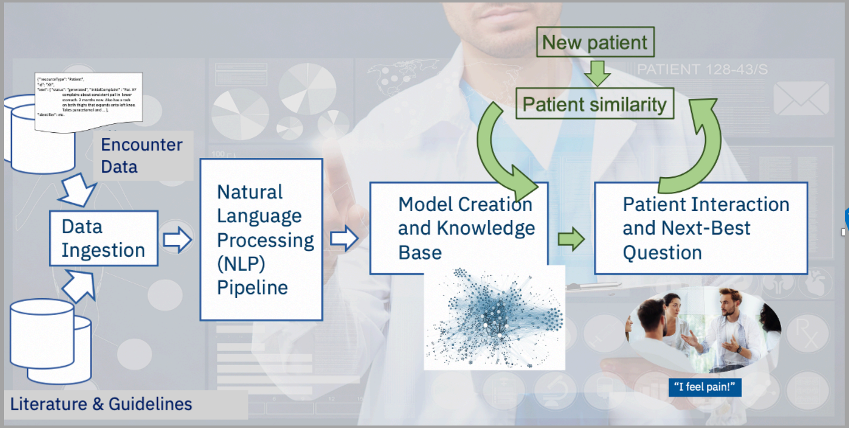
- Behavior phenotype-based care planning strategies could yield more effective intervention recommendations for goal attainment compared to population level strategies.



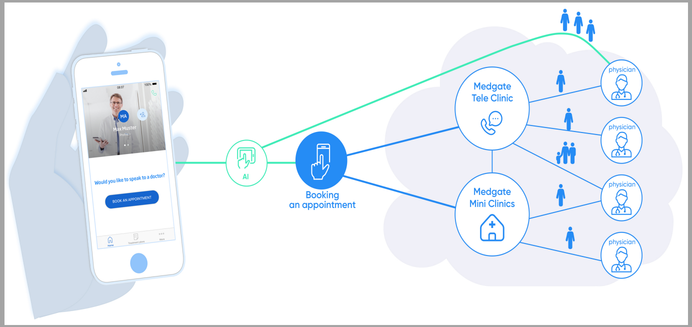
AMIA 2017, AMIA 2018 (best paper)

AI Triage Engine

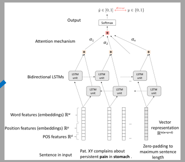
Decision support system for medical triage to guide individuals to the next step of care



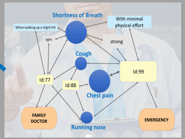
AIDA



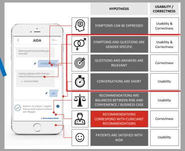
Medical triage tool deployed by a telemedicine service provider in Switzerland



NLP Pipeline (CNN, Bi-GRU, Bi-LSTM)



Auto-creation of ontology and language agnostic KG

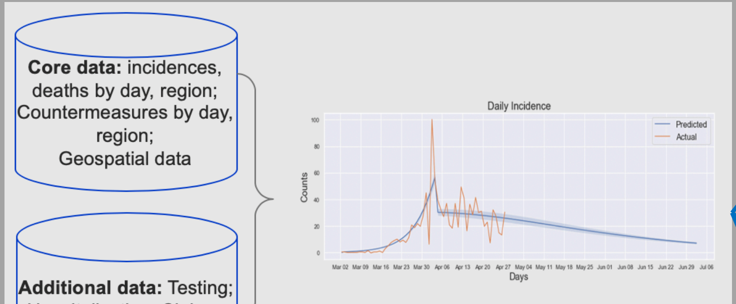


ML models for next-best-question and recommendation

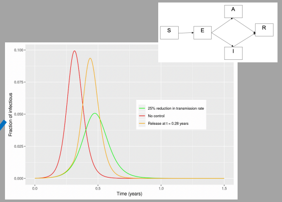
EMNLP 2018, EMBS 2018, AMIA 2020

Hyperlocal Case Prediction

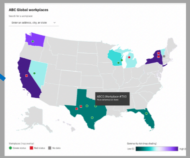
Framework for hyperlocal predictions of COVID-19 cases using novel compartmental models with ML enhancements



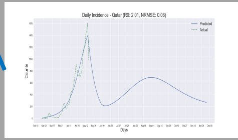
County/State/Country level daily prediction of cases with uncertainty under current or what-if conditions



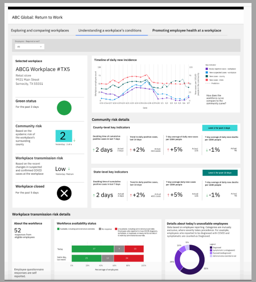
Compartmental model (SEAIR) accounting for asymptomatic transmission, NPI and testing



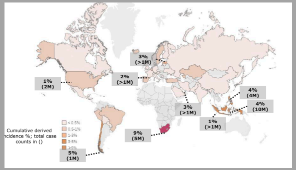
Spatial Temporal Clustering



Detection of Non-Pharmaceutical Interventions



Community risk model for Return To Work Advisor



Vaccine trial site identification

Resource Demand Prediction

Health Security 2019, MedRxiv 2020, KDD epiDAMIK 2020

Thank You !