# Predictive Modeling with Longitudinal Patient Clinical Records

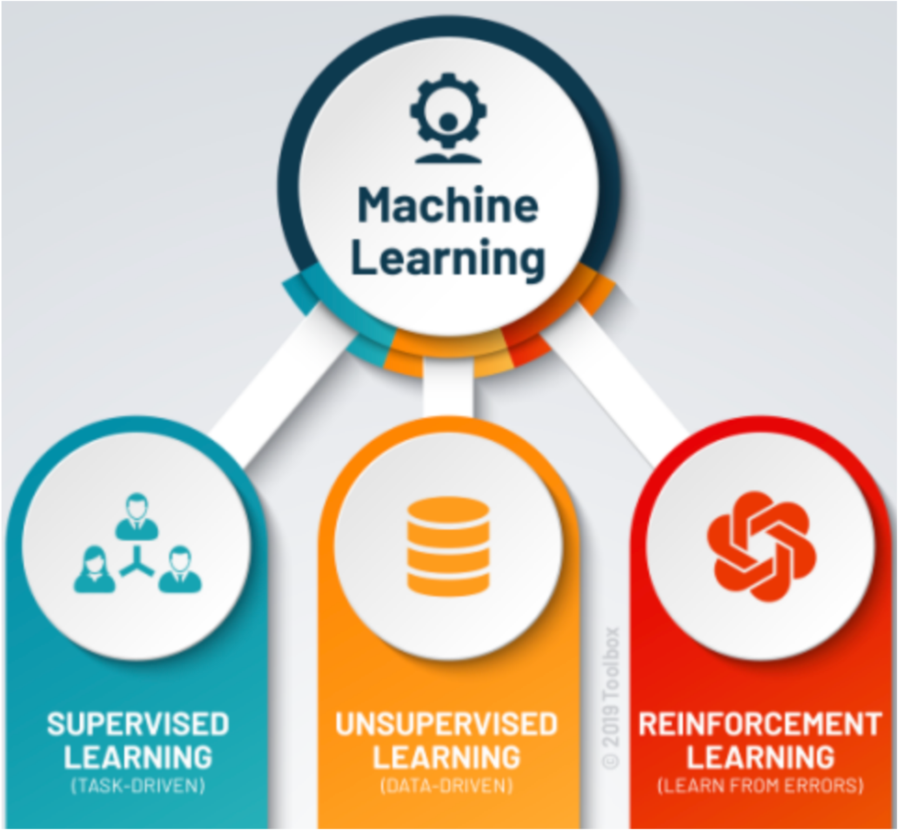

Fei Wang
Associate Professor
Department of Population Health Sciences
Weill Cornell Medicine
feiwang.cornell@gmail.com
@feiwang03
https://wcm-wanglab.github.io/index.html

# Disclosures

- I am consulting for the following companies
  - IBM
  - Boehringer Ingelheim
  - American Air Liquide

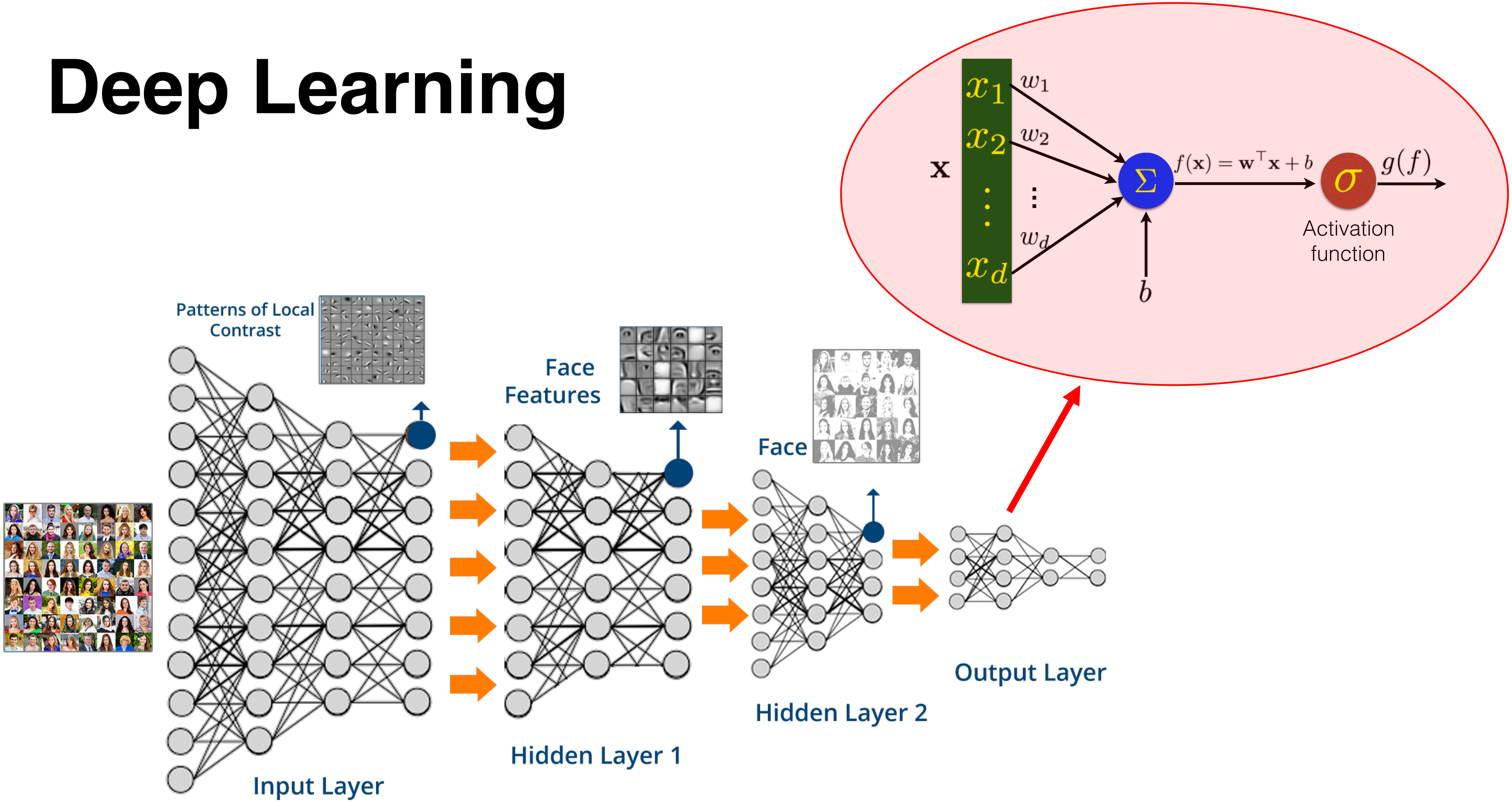
- Funding acknowledgement at the end

# Machine Learning




https://www.potentiaco.com/what-is-machine-learning-definition-types-applications-and-examples/

# Deep Learning




https://www.michaelchimenti.com/2017/11/deep-neural-nets-software-2-0/

# Deep Learning

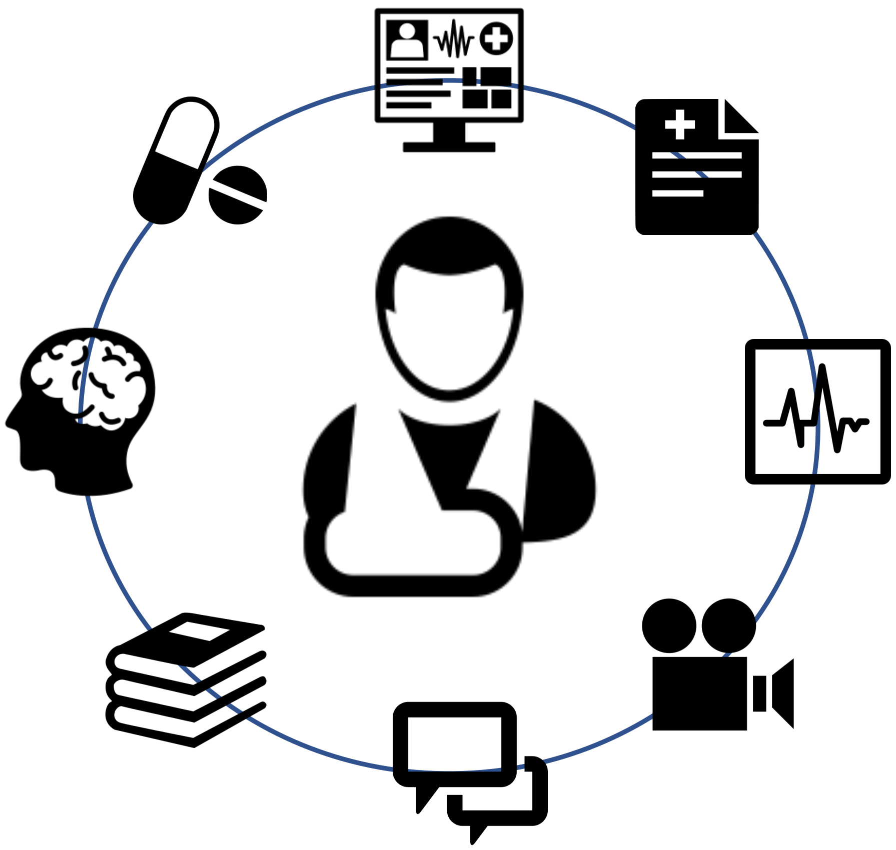
https://qbi.uq.edu.au/blog/2017/10/google-alphago-zero-masters-game-three-days


https://electrek.co/2017/04/29/elon-musk-tesla-plan-level-5-full-autonomous-driving/


https://www.shellypalmer.com/2017/01/5-awesome-illegal-uses-alexa/

https://siliconangle.com/2020/07/19/openais-latest-ai-text-generator-gpt-3-amazes-early-adopters/

# Medicine



The NEW ENGLAND JOURNAL of MEDICINE

Perspective

A New Initiative on Precision Medicine

Francis S. Collins, M.D., Ph.D., and Harold Varmus, M.D.
N Engl J Med 2015; 372:793-795 | February 26, 2015 | DOI: 10.1056/NEJMp1500523

Comments open through March 4, 2015

Share:

Article | References | Citing Articles (784) | Comments (7) | Metrics

"Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes — and to give all of us access to the personalized information we need to keep ourselves and our families healthier."

— President Barack Obama, State of the Union Address, January 20, 2015

President Obama has long expressed a strong conviction that science offers great potential for improving health. Now, the President has announced a research initiative that aims to accelerate progress toward a new era of precision medicine (www.whitehouse.gov/precisionmedicine). We believe that the time is right for this visionary initiative, and the National Institutes of Health (NIH) and other partners will work to achieve this vision.

AUDIO INTERVIEW

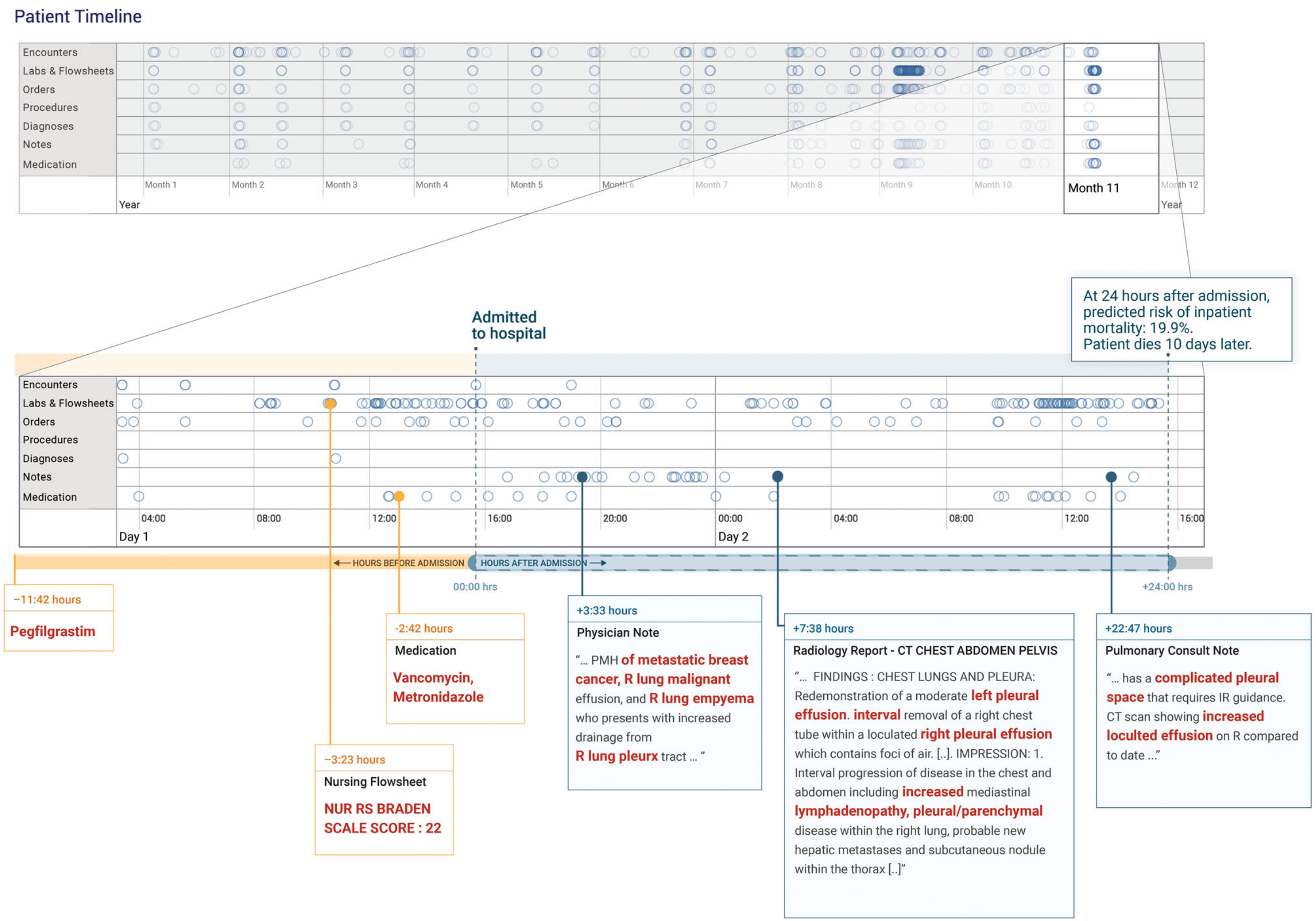
Interview with Dr. Francis Collins on what to expect from the recently announced Precision Medicine Initiative. (10:07)

Listen

Download

"The initiative will encourage and support the next generation of scientists to develop creative new approaches for detecting, measuring, and analyzing a wide range of biomedical information — including molecular, genomic, cellular, clinical, behavioral, physiological, and environmental parameters"

# Electronic Health Records

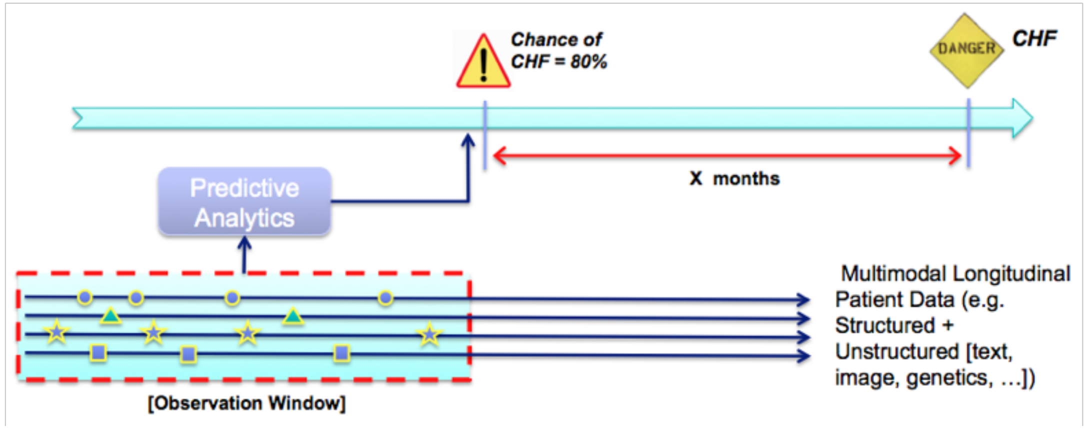



https://www.nature.com/articles/s41746-018-0029-1

# The Journey

# Risk Prediction

- **_Goal_**:
  - Build a model for predicting HF onset x months before the HF diagnosis
- **_Data_: Longitudinal patient records**
  - Structured data:
    - Demographics, Outpatient diagnoses, Problem List , Vitals, Medication, Labs
  - Unstructured text : encounter notes
- **_Challenge_ faced by our clinical partners:**
  - How to systematically collect and evaluate many weak and non-specific indicators and identify the ones that combined are truly predictive




- Incorporating knowledge driven risk factors
- Accurate prediction
- Minimal redundancy:
  - Little correlation between the selected data driven risk factors and existing knowledge driven risk factor
  - Little correlation among the additional risk factors from data, to further ensure quality of the additional factors

$$f(\boldsymbol{\alpha}) = \underbrace{\boxed{\mathcal{L}(\mathbf{y}, \mathbf{X}\boldsymbol{\alpha})}}_{\text{Model Accuracy}} + \frac{\beta}{4}\left[\underbrace{\boxed{\sum_{ij\in\mathcal{D}} (\alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j)^2}}_{\text{Correlation between data-driven features}} + \underbrace{\boxed{\sum_{i\in\mathcal{D}, j\in\mathcal{K}} (\alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j)^2}}_{\text{Correlation between data- and knowledge-driven features}}\right] + \lambda \underbrace{\boxed{\|\boldsymbol{\alpha}\|_1}}_{\text{Sparsity Penalty}}$$

| Category |
|---|
| Diagnosis |
| Medication |
| Lab |
| Symptom |

| Feature | Relevancy to HF |
|---|---|
| Dyslipidemia | ✓ |
| Thiazides-like Diuretics | ✓ |
| Antihypertensive Combinations | ✓ |
| Aminopenicillins | ✓ |
| Bone density regulators | ✗ |
| Naturietic Peptide | ✓ |
| Rales | ✓ |
| Diuretic Combinations | ✓ |
| S3Gallop | ✓ |
| NSAIDS | ✓ |

Sun, Jimeng, Jianying Hu, Dijun Luo, Marianthi Markatou, **Fei Wang**, Shahram Edabollahi, Steven E. Steinhubl, Zahra Daar, and Walter F. Stewart. "Combining knowledge and data driven insights for identifying risk factors using electronic health records." In *AMIA Annual Symposium Proceedings*, vol. 2012, p. 901. American Medical Informatics Association, 2012.

Proceedings of the 2012 SIAM International Conference on Data Mining

< Previous Chapter | Table of Contents | Next Chapter >

Abstract | PDF

## SOR: Scalable Orthogonal Regression for Non-Redundant Feature Selection and its Healthcare Applications

Dijun Luo, Fei Wang, Jimeng Sun, Marianthi Markatou, Jianying Hu and Shahram Ebadollahi

This Paper Appears in

2012 SIAM International Conference on DATA MINING

Abstract

As more clinical information with increasing diversity become available for analysis, a large number of features can be constructed and leveraged for predictive modeling. Feature selection is a classic analytic component that faces new challenges due to the new

> AMIA Annu Symp Proc. 2012;2012:901-10. Epub 2012 Nov 3.

## Combining knowledge and data driven insights for identifying risk factors using electronic health records

Jimeng Sun [1], Jianying Hu, Dijun Luo, Marianthi Markatou, Fei Wang, Shahram Edabollahi, Steven E Steinhubl, Zahra Daar, Walter F Stewart

Affiliations + expand

PMID: 23304365 PMCID: PMC3540578

Free PMC article

JACC: HEART FAILURE VOL. 2, NO. 4, 2014
© 2014 BY THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION ISSN 2213-1779/$36.00
PUBLISHED BY ELSEVIER INC. http://dx.doi.org/10.1016/j.jchf.2014.03.010

## Low Bone Mineral Density Predicts Incident Heart Failure in Men and Women

CME

### The EPIC (European Prospective Investigation Into Cancer and Nutrition)-Norfolk Prospective Study

Roman Pfister, MD,* Guido Michels, MD,* Stephen J. Sharp, MSc,† Robert Luben, BSc,‡
Nick J. Wareham, MBBS, PhD,† Kay-Tee Khaw, MBBChir, PhD‡

AMJ AME MEDICAL JOURNAL
AN OPEN ACCESS GENERAL MEDICAL JOURNAL

Home | Journal Info | For Authors | For Reviewers | Special Contents | All Articles | Archives | News

Home / March 2018 / Cardiovascular disease and bone loss—new research in identifying common disease pathophysiologies and predictors

Editorial

Check for up

## Cardiovascular disease and bone loss—new research in identifying common disease pathophysiologies and predictors

Sarah L. West[1], Emma O'Donnell[2]

JAHA
Journal of the American Heart Association

AHA Journals | Journal Information | All Issues | Subjects | Features | Resour

Home > Journal of the American Heart Association > Vol. 6, No. 3 > Bone Mineral Density and Risk of Heart Failure in ...

OPEN ACCESS
ARTICLE

Download PDF

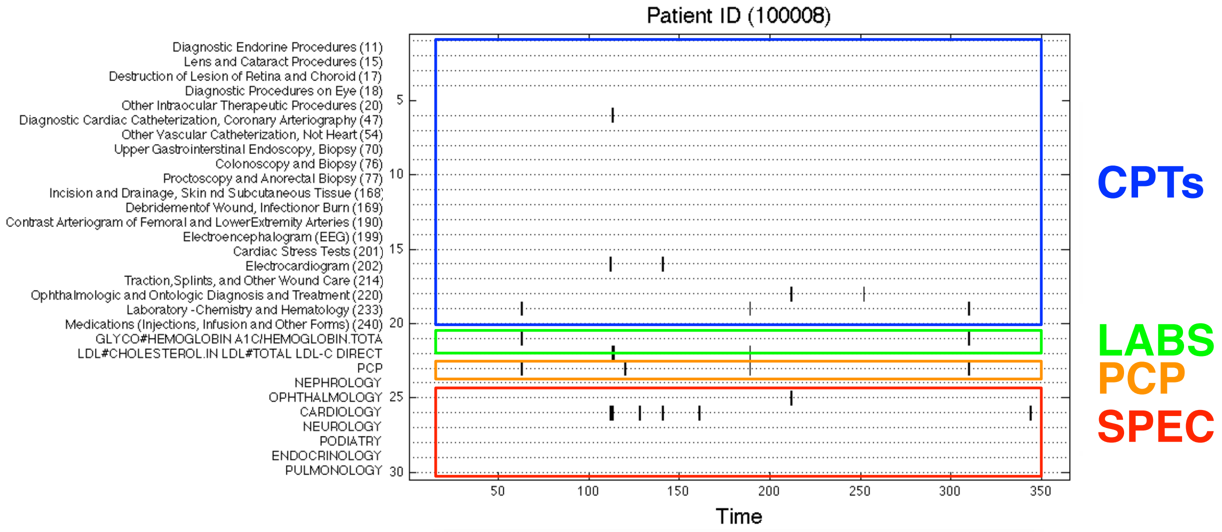
Tools Share

## Bone Mineral Density and Risk of Heart Failure in Older Adults: The Cardiovascular Health Study

Raymond B. Fohtung, David L. Brown, William J. H. Koh, Traci M. Bartz, Laura D. Carbone, Roberto Civitelli, Phyllis K. Stein, Paulo H. M. Chaves, Bryan R. Kestenbaum, and Jorge R. Kizer

Originally published 13 Mar 2017 | https://doi.org/10.1161/JAHA.116.004344 | Journal of the American Heart Association. ;6:e004344

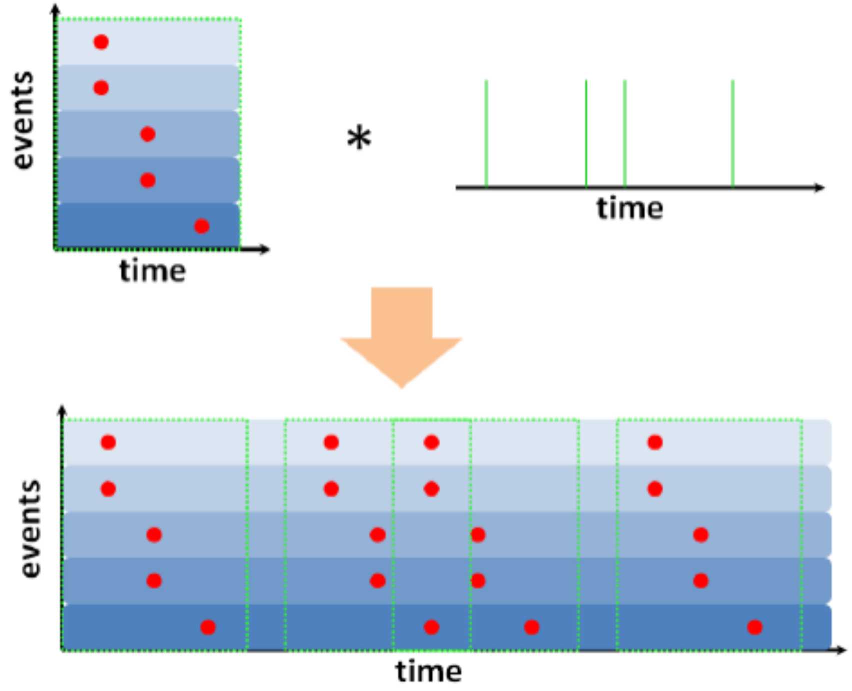
# Matrix Representation

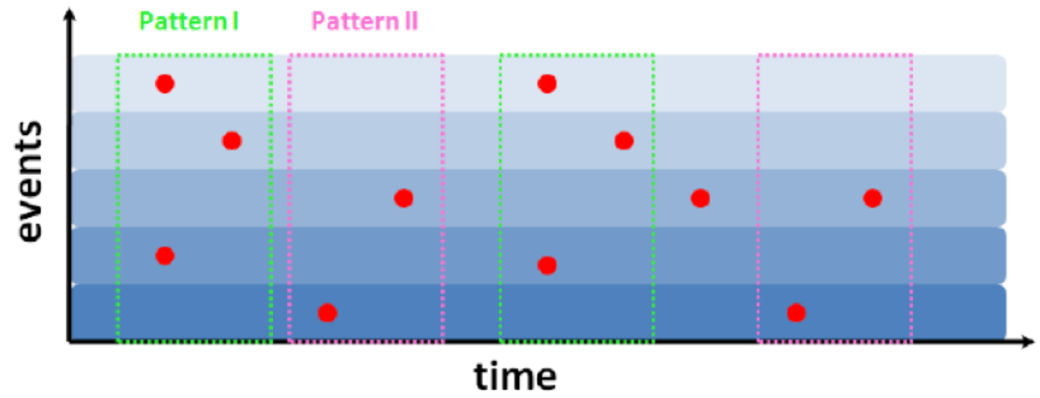



**Definition** (One-Side Convolution). *The one-side convolution of* $\mathbf{F} \in \mathbb{R}^{n\times m}$ *and* $\mathbf{g} \in \mathbb{R}^{t\times 1}$ *is an* $n \times t$ *matrix with*

$$(\mathbf{F} * \mathbf{g})_{ij} = \sum\nolimits_{k=1}^{t} g_{j-k+1} F_{ik}$$

*Note that* $g_j = 0$ *if* $j \leqslant 0$ *or* $j > t$, *and* $F_{ik} = 0$ *if* $k > m$.



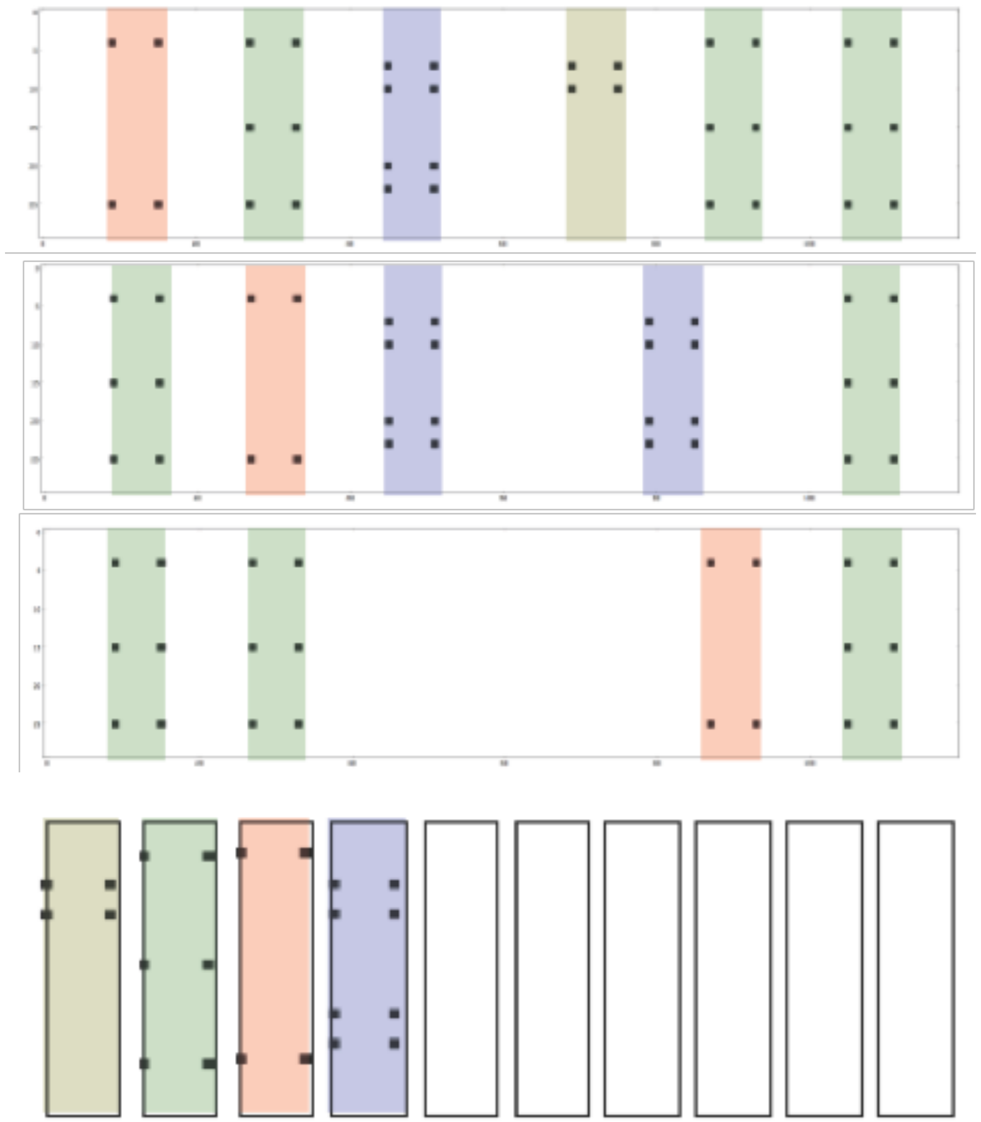

$$\min_{\mathcal{F},\mathcal{G}} \quad \mathcal{J}$$

$$s.t. \quad \forall r = 1, \cdots, R; c = 1, \cdots, C$$

$$\mathbf{F}^{(r)} \geqslant 0, \mathbf{g}_c^{(r)} \geqslant 0$$

$$\mathcal{J} = \sum_{c=1}^{C} d_\beta \left( \mathbf{A}_c \odot \mathbf{X}_c, \mathbf{A}_c \odot \left( \sum_{r=1}^{R} \mathbf{F}^{(r)} * \mathbf{g}_c^{(r)} \right) \right) + \lambda_1 \sum_{r=1}^{R} \|\mathbf{F}^{(r)}\|_1 + \lambda_2 \sum_{c=1}^{C} \sum_{r=1}^{R} \|\mathbf{g}_c^{(r)}\|_1$$


**Wang, Fei**, Noah Lee, Jianying Hu, Jimeng Sun, and Shahram Ebadollahi. "Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach." In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 453-461. 2012.



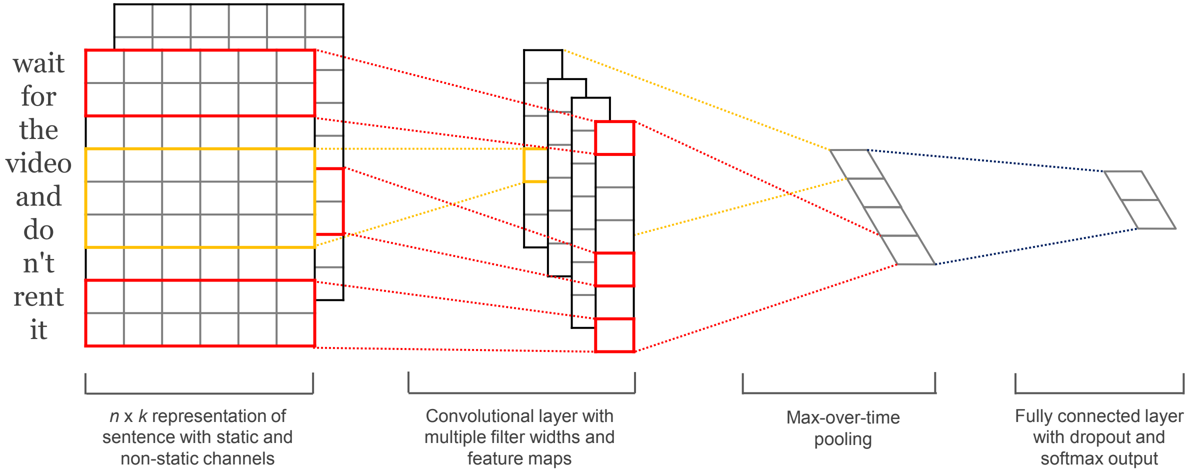
Wang, Fei, Noah Lee, Jianying Hu, Jimeng Sun, Shahram Ebadollahi, and Andrew F. Laine. "A framework for mining signatures from event sequences and its applications in healthcare data." *IEEE transactions on pattern analysis and machine intelligence* 35, no. 2 (2012): 272-285.

EHR of a pool of diabetic patients (n = 21,384)

| CPT Code | $G_1$ Description |
|---|---|
| 11 | Diagnostic Endorine Procedures |
| 15 | Lens and Cataract Procedures |
| 17 | Destruction of Lesion of Retina and Choroid |
| 18 | Diagnostic Procedures on Eye |
| 20 | Other Intraocular Therapeutic Procedures |
| 47 | Diagnostic Cardiac Catheterization, Coronary Arteriography |
| 54 | Other Vascular Catheterization, Not Heart |
| 70 | Upper Gastrointerstinal Endoscopy, Biopsy |
| 76 | Colonoscopy and Biopsy |
| 77 | Proctoscopy and Anorectal Biopsy |
| 168 | Incision and Drainage, Skin nd Subcutaneous Tissue |
| 169 | Debridementof Wound, Infectionor Burn |
| 190 | Contrast Arteriogram of Femoral and LowerExtremity Arteries |
| 199 | Electroencephalogram (EEG) |
| 201 | Cardiac Stress Tests |
| 202 | Electrocardiogram |
| 214 | Traction,Splints, and Other Wound Care |
| 220 | Ophthalmologic and Ontologic Diagnosis and Treatment |
| 233 | Laboratory -Chemistry and Hematology |
| 240 | Medications (Injections, Infusion and Other Forms) |

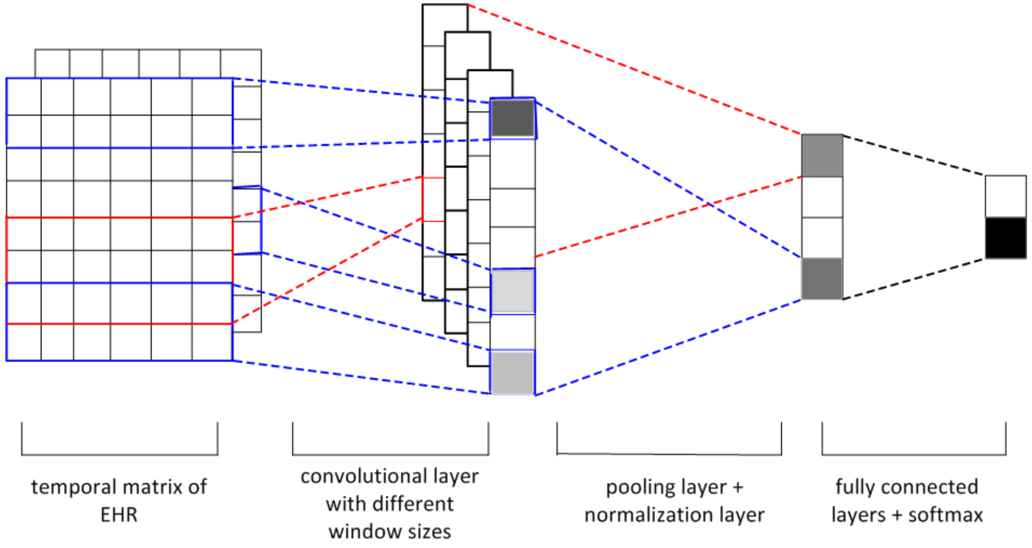
| LABS | $G_2$ Description |
|---|---|
| | GLYCO and HEMOGLOBIN A1C/HEMOGLOBIN.TOTA |
| | LDL, CHOLESTEROL.IN LDL, and TOTAL LDL-C DIRECT |
| **PCP** | **$G_3$ Description** |
| | General Primary Care Physician Visits |
| **SPECIALTY** | **$G_4$ Description** |
| | NEPHROLOGY |
| | OPHTHALMOLOGY |
| | CARDIOLOGY |
| | NEUROLOGY |
| | PODIATRY |
| | ENDOCRINOLOGY |
| | PULMONOLOGY |

- Repeated high Hemoglobin A1C value
- Repeated cardiac disease related procedure
- Repeated lab test (CPT code 233)
- Co-occurrence of high Hemoglobin A1C value and high Cholesterol

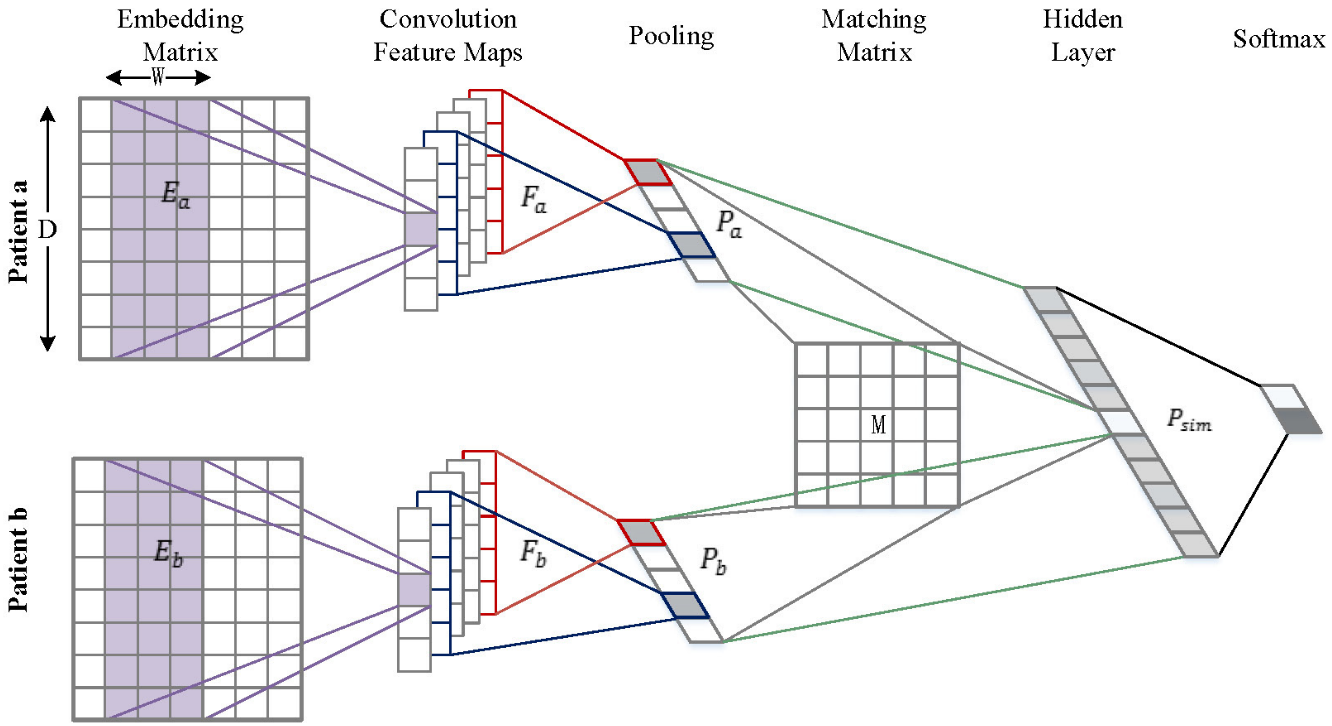



Kim, Yoon. "Convolutional Neural Networks for Sentence Classification." ENMLP 2014.

# CNN for EHR Analysis

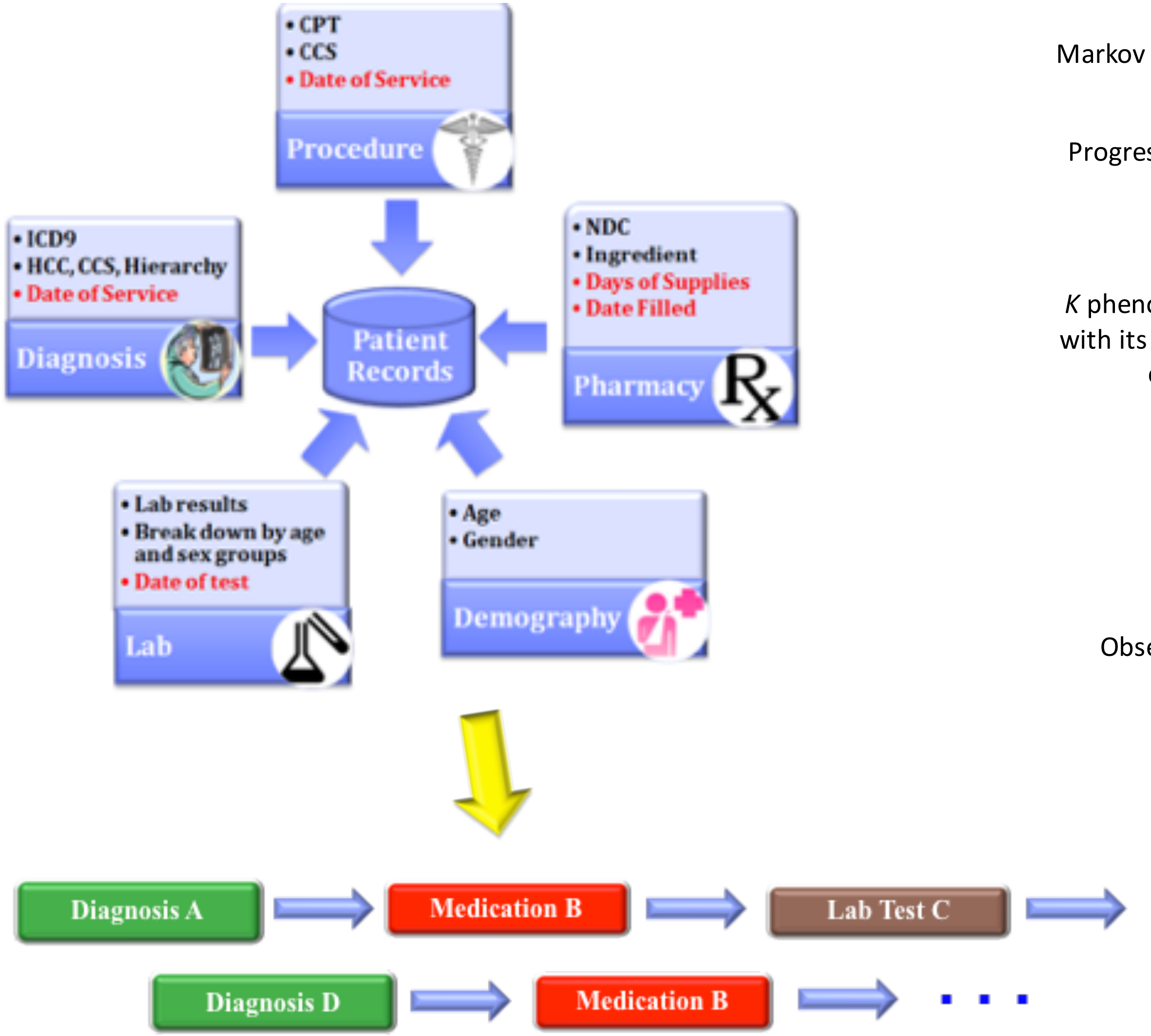


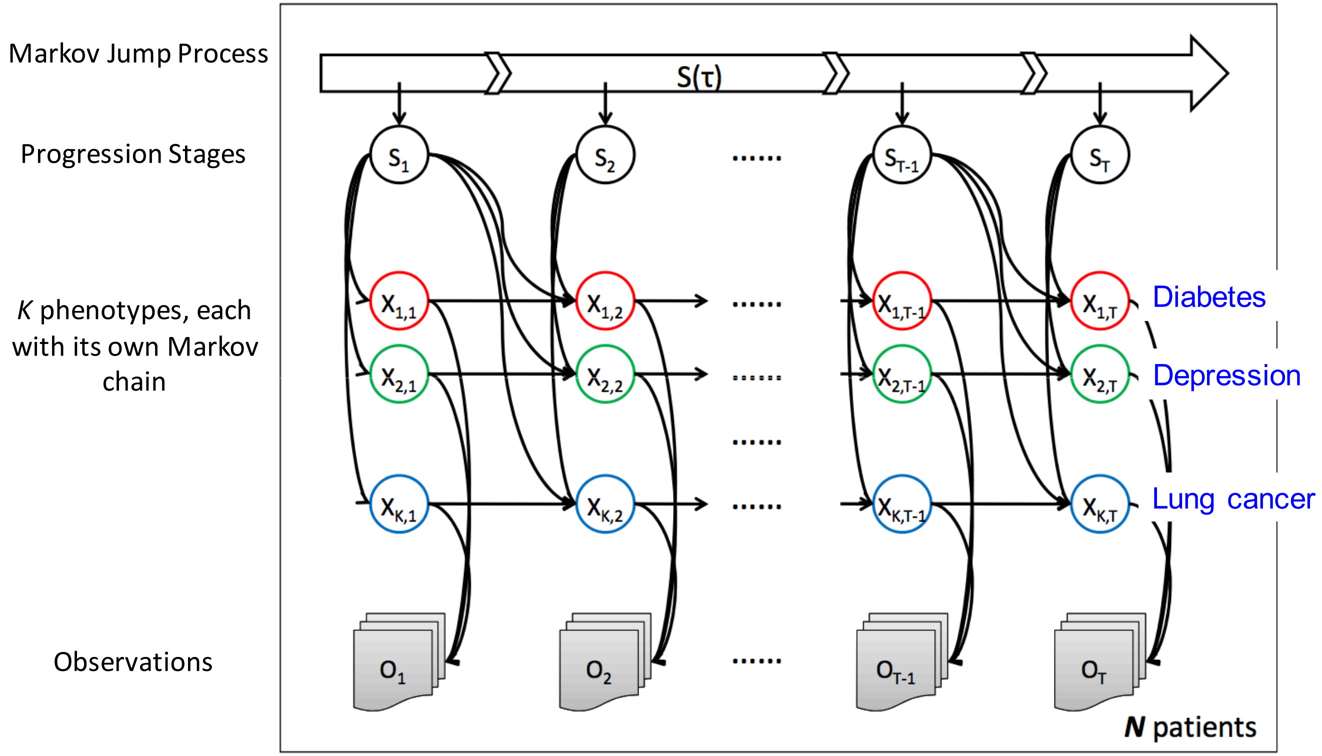

Cheng, Yu, **Fei Wang**, Ping Zhang, and Jianying Hu. "Risk prediction with electronic health records: A deep learning approach." In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pp. 432-440. Society for Industrial and Applied Mathematics, 2016.




Zhu, Zihao, Changchang Yin, Buyue Qian, Yu Cheng, Jishang Wei, and **Fei Wang**. "Measuring patient similarities via a deep architecture with medical concept embedding." In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 749-758. IEEE, 2016.

# Sequence Representation




Wang, Xiang, David Sontag, and **Fei Wang**. "Unsupervised learning of disease progression models." *In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 85-94. 2014.

# Progression Subtyping

THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

## 2016 PPMI DATA CHALLENGE

The Parkinson's Progression Markers Initiative (PPMI) has generated a comprehensive, standardized, longitudinal set of clinical, biological and imaging data unique to the Parkinson's disease (PD) field and ripe for novel and innovative exploration. In mid-2016, study sponsor The Michael J. Fox Foundation (MJFF) cast an open call to computational scientists, data scientists and neuroscientists to analyze PPMI data toward new insights into PD diagnosis and progression.

Parkinson's is highly variable, with age of onset, rate of progression, and type and severity of symptoms different across the 5 million worldwide living with the disease. Identifying models for prognosis and sub-typing would aid in subject selection for clinical studies and design of trials toward novel therapies.

PPMI data is uploaded in real time and accessible to qualified researchers. Learn more and download the data by visiting www.ppmi-info.org.

1. What factors at baseline predict clinical progression?
2. What are the sub–types of Parkinson's disease?

## UNIVERSITY OF CALIFORNIA, SAN FRANCISCO AND WEILL CORNELL MEDICINE RESEARCHERS NAMED WINNERS OF 2016 PARKINSON'S PROGRESSION MARKERS INITIATIVE DATA CHALLENGE

November 29,2016

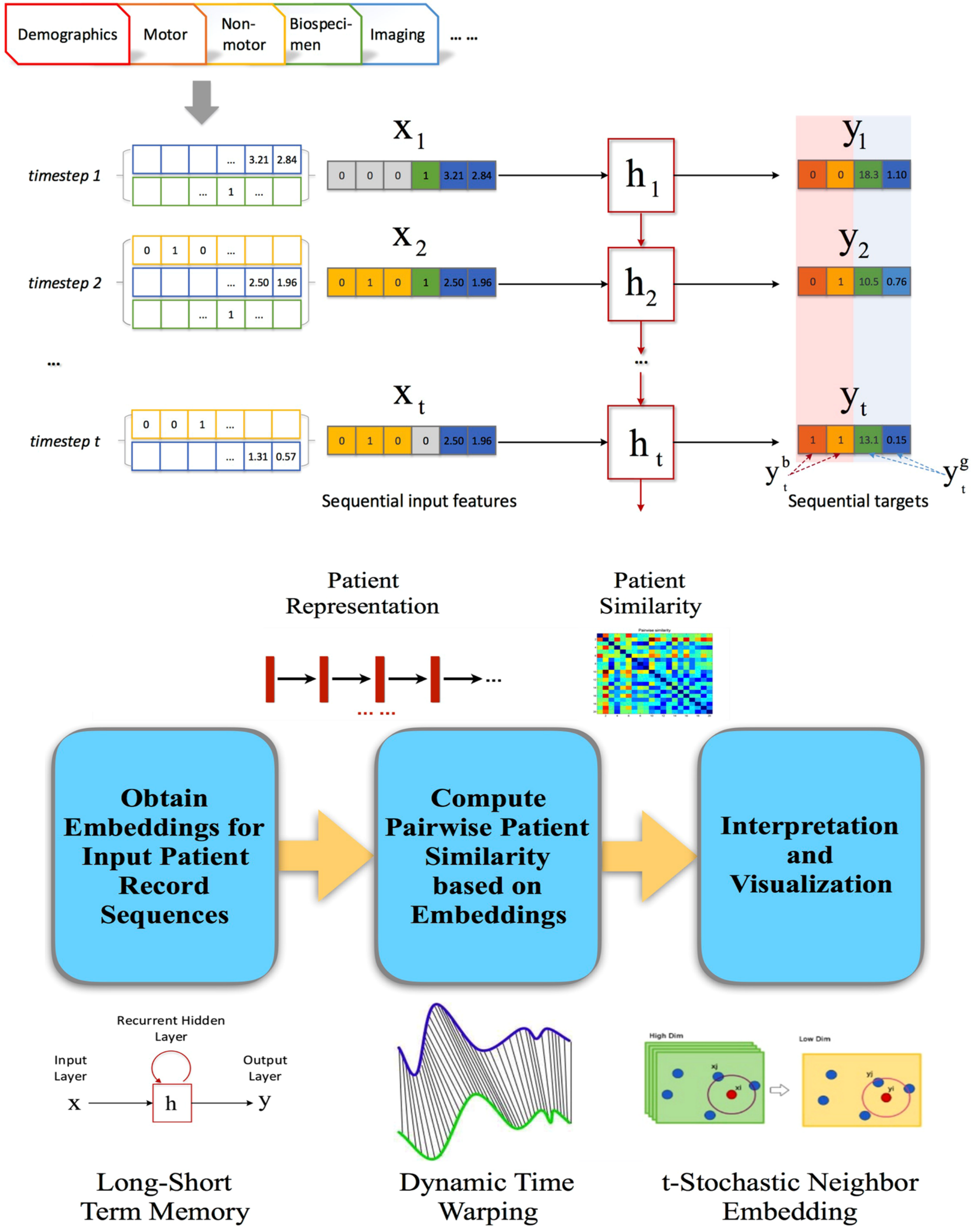
***The Michael J. Fox Foundation Hosted the Challenge toward Computational Analysis of the Robust Study Dataset***

***Answering Fundamental Questions on Parkinson's Progression and Subtypes Will Assist in Development and Testing of New Therapies***

***Each Winner Receives a $25,000 Award***

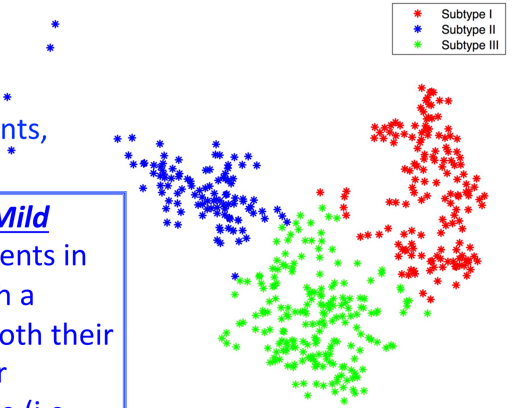
**NEW YORK and SAN FRANCISCO** -- The Michael J. Fox Foundation for Parkinson's Research (MJFF) announces Duygu Tosun-Turgut, PhD, assistant professor of radiology and biomedical imaging at UC San Francisco and co-director of the Center for Imaging of Neurodegenerative Diseases at the San Francisco Veterans Affairs Health Care System; and Fei Wang, PhD, assistant professor of health care policy and research at Weill Cornell Medicine as winners of the MJFF-led 2016 Parkinson's Progression Markers Initiative (PPMI) Data Challenge. Each will receive a $25,000 award furnished by MJFF and supported in part by GE Healthcare.

https://www.michaeljfox.org/foundation/publication-detail.html?id=625&category=7&j=737845)




Subtype II (107 patients, avg age 61.93)

***Moderate Baseline, Mild Progression***: The patients in this subtype start with a moderate status on both their motor and non-motor capabilities at baseline (i.e., more severe than Subtype I). Both their motor and non-motor functionalities progress slowly over time.



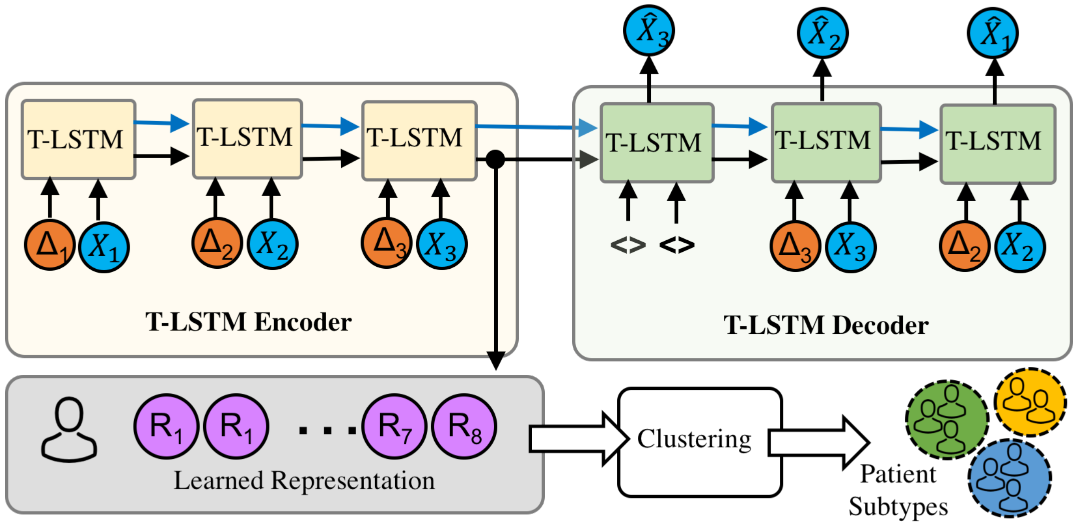

Subtype I (201 patients, avg age 58.79)

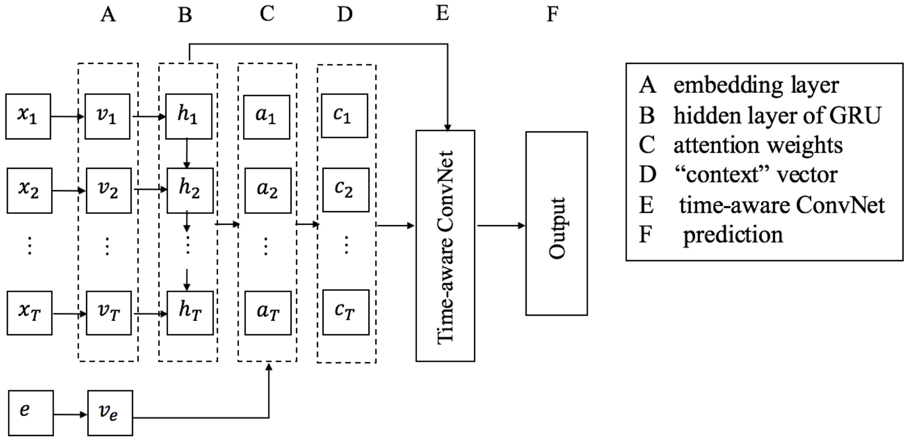
***Mild Baseline, Moderate Motor Progression***: The patients in this subtype start with a mild status on both their motor and non-motor capabilities at baseline. However, their motor functionalities will decay at a moderate rate over time while their cognitive abilities are relatively stable.

Subtype III (158 patients, avg age 65.32)

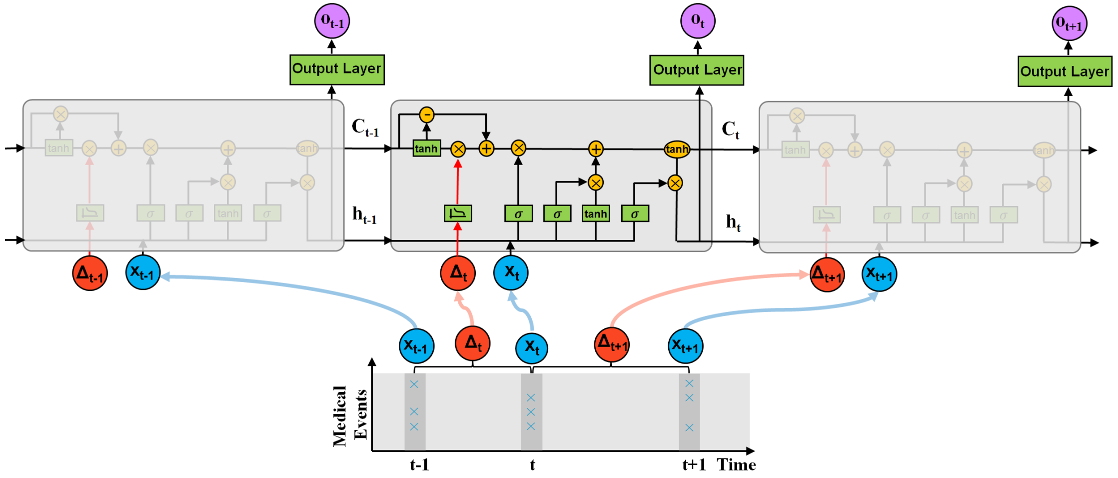
***Severe Baseline, Rapid Progression***: The patients in this subtype start with a severe status on both their motor and non-motor capabilities at baseline (i.e., more severe than Subtype I and II). Both their motor and non-motor functionalities progress rapidly over time.


Zhang, Xi, Jingyuan Chou, Jian Liang, Cao Xiao, Yize Zhao, Harini Sarva, Claire Henchcliffe, and **Fei Wang**. "Data-driven subtyping of Parkinson's disease using longitudinal clinical records: a cohort study." *Scientific reports* 9, no. 1 (2019): 1-12.

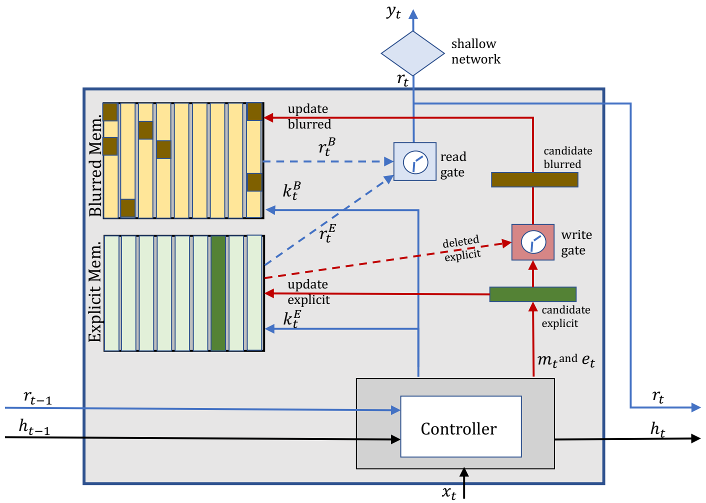



Ma, Tengfei, Cao Xiao, and **Fei Wang**. "Health-atm: A deep architecture for multifaceted patient health record representation and risk prediction." In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pp. 261-269. Society for Industrial and Applied Mathematics, 2018.




Baytas, Inci M., Cao Xiao, Xi Zhang, **Fei Wang**, Anil K. Jain, and Jiayu Zhou. "Patient subtyping via time-aware LSTM networks." In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 65-74. 2017.




Chakraborty, Prithwish, **Fei Wang**, Jianying Hu, and Daby Sow. "Explicit-Blurred Memory Network for Analyzing Patient Electronic Health Records." *arXiv preprint arXiv:1911.06472* (2019).

# The Challenges

npj | digital medicine

Explore our content ⌄ Journal information ⌄

nature > npj digital medicine > articles > article

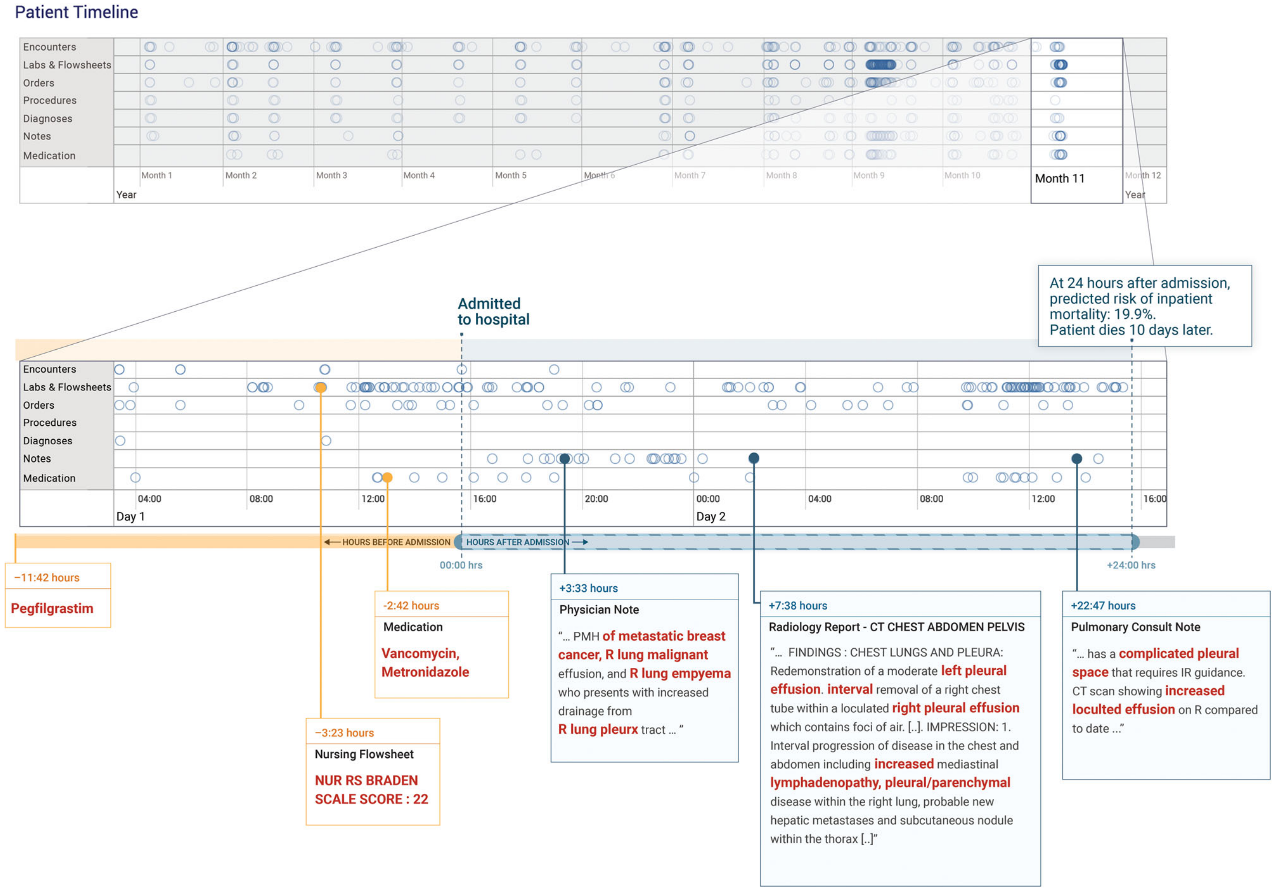
Article | Open Access | Published: 08 May 2018

# Scalable and accurate deep learning with electronic health records

Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, Patrik Sundberg, Hector Yee, Kun Zhang, Yi Zhang, Gerardo Flores, Gavin E. Duggan, Jamie Irvine, Quoc Le, Kurt Litsch, Alexander Mossin, Justin Tansuwan, De Wang, James Wexler, Jimbo Wilson, Dana Ludwig, Samuel L. Volchenboum, Katherine Chou, Michael Pearson, Srinivasan Madabushi, Nigam H. Shah, Atul J. Butte, Michael D. Howell, Claire Cui, Greg S. Corrado & Jeffrey Dean -Show fewer authors

*npj Digital Medicine* **1**, Article number: 18 (2018) | Cite this article






https://www.nature.com/articles/s41746-018-0029-1

[HTML] **Scalable and accurate deep learning with electronic health records**

A Rajkomar, E Oren, K Chen, AM Dai, N Hajaj… - NPJ Digital …, 2018 - nature.com

Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction of curated predictor variables from normalized EHR …

☆ 99 Cited by 802 Related articles All 16 versions

## In-hospital mortality prediction

- the most recent systolic blood pressure, heart-rate, respiratory rate and temperature in Fahrenheit
- the most recent white blood cell count hemoglobin, sodium, creatinine, troponin, lactate oxygen saturation, oxygen source, glucose, calcium, potassium, chloride, blood urea nitrogen (BUN), carbon dioxide, hematocrit, platelet, magnesium, phosphorus, albumin, aspartate transaminase (AST), Alkaline Phosphatase, Total Bilirubin, International Normalized Ratio, and Absolute Neutrophil Count (ANC)

## Hospital readmission prediction

| | | | |
|---|---|---|---|
| Hemoglobin level at discharge, g/dL | ≥12 0 | <12 +1 | |
| Discharged from oncology service | No 0 | Yes +2 | |
| Sodium level at discharge, mEq/L | ≥135 0 | <135 +1 | |
| Any ICD-9 coded procedure performed during hospital stay | No 0 | Yes +1 | |
| Index admission type | Elective 0 | Urgent or emergent +1 | |
| Number of hospital admissions during the previous year | 0-1 0 | 2-5 +2 | >5 +5 |
| Length of stay | <5 days 0 | ≥5 days +2 | |

## Length-of-stay prediction

- prior HCC codes in the timeline (counts for each one), the principal diagnosis coded as a CCS, hospital service, and the most recent lab value of each possible lab used in the mortality baseline model.

**Weighted RNN**

**Feedforward NN with Time-Aware Attention**

**Boosted Decision Stumps**

| | Hospital A | Hospital B |
|---|---|---|
| **Inpatient Mortality, AUROC[1](95% CI)** | | |
| Deep learning 24 hours after admission | **0.95**(0.94-0.96) | **0.93**(0.92-0.94) |
| Full feature enhanced baseline at 24 hours after admission | 0.93 (0.92-0.95) | 0.91 (0.89-0.92) |
| Full feature simple baseline at 24 hours after admission | 0.93 (0.91-0.94) | 0.90 (0.88-0.92) |
| Baseline (aEWS[2]) at 24 hours after admission | 0.85 (0.81-0.89) | 0.86 (0.83-0.88) |
| **30-day Readmission, AUROC (95% CI)** | | |
| Deep learning at discharge | **0.77**(0.75-0.78) | **0.76**(0.75-0.77) |
| Full feature enhanced baseline at discharge | 0.75 (0.73-0.76) | 0.75 (0.74-0.76) |
| Full feature simple baseline at discharge | 0.74 (0.73-0.76) | 0.73 (0.72-0.74) |
| Baseline (mHOSPITAL[3]) at discharge | 0.70 (0.68-0.72) | 0.68 (0.67-0.69) |
| **Length of Stay at least 7 days AUROC (95% CI)** | | |
| Deep learning 24 hours after admission | **0.86**(0.86-0.87) | **0.85**(0.85-0.86) |
| Full feature enhanced baseline at 24 hours after admission | 0.85 (0.84-0.85) | 0.83 (0.83-0.84) |
| Full feature simple baseline at 24 hours after admission | 0.83 (0.82-0.84) | 0.81 (0.80-0.82) |
| Baseline (mLiu[4]) at 24 hours after admission | 0.76 (0.75-0.77) | 0.74 (0.73-0.75) |

[1] Area under the receiver operator curve
[2] Augmented early warning score
[3] Modified HOSPITAL score
[4] Modified Liu score

## Editor's choice: machine learning in healthcare

**scientific** reports

**Explore our content** ⌄ **Journal information** ⌄

nature > scientific reports > articles > article

Article | Open Access | Published: 20 February 2019

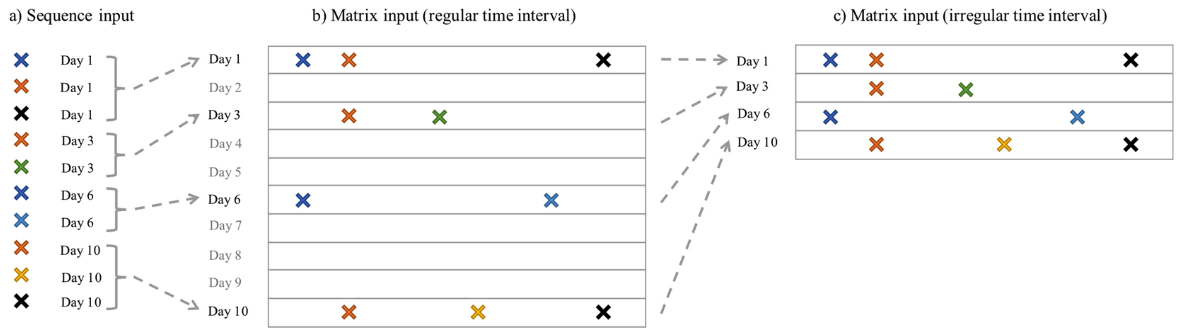
# Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD

Xu Min, Bin Yu & Fei Wang

*Scientific Reports* **9**, Article number: 2362 (2019) | Cite this article

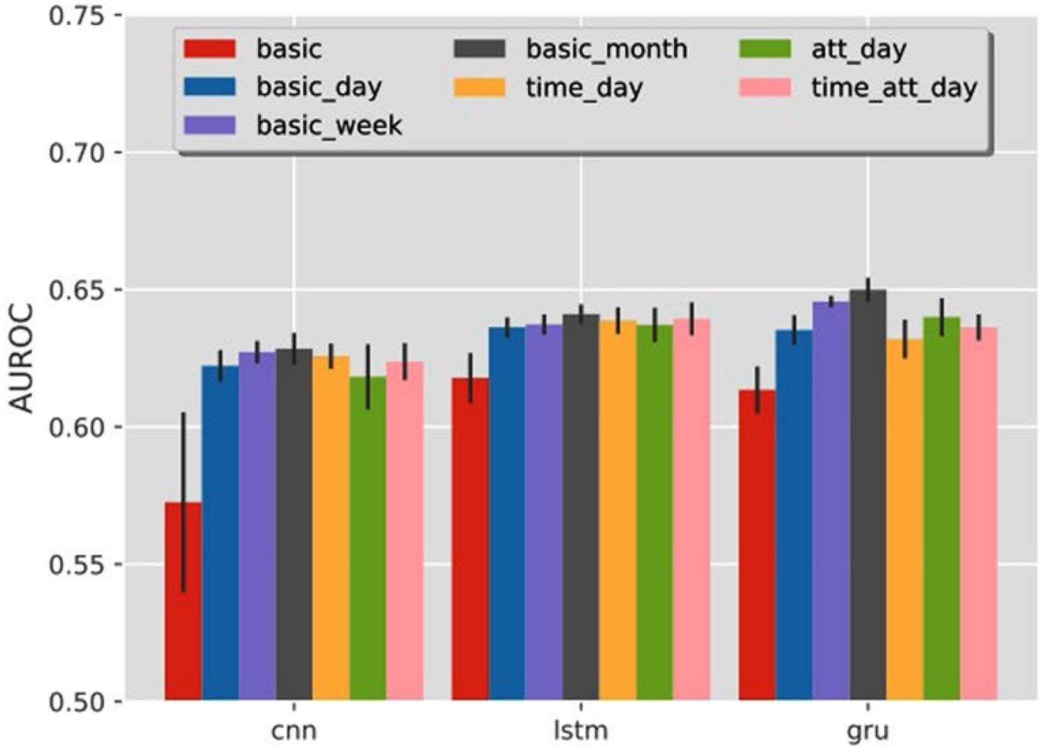


https://www.nature.com/articles/s41598-019-39071-y

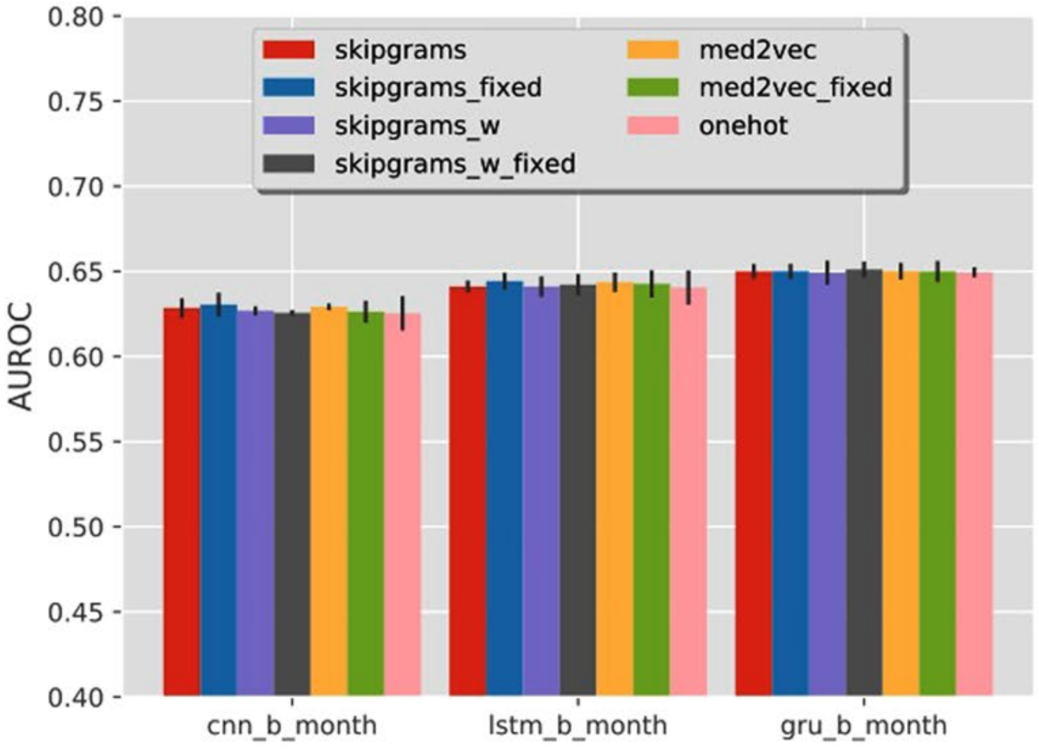



| **Feature x** | | **Dimension (one-year history)** | **Dimension (full history)** |
|---|---|---|---|
| Knowledge-driven | HOS | 4 | — |
| | LACE | 4 | — |
| | hand | 12 | 12 |
| Data-driven | DX | 9743 | 10306 |
| | DX_3dig | 1153 | 1169 |
| | DX_CCS | 285 | 285 |
| | DX_HCC | 197 | 197 |
| | PROC | 11193 | 12009 |
| | PROC_group | 399 | 402 |
| | PHAR | 20289 | 22964 |
| | PHAR_GTC | 42 | 42 |
| | LC | 32 | 33 |

| | LR | LR_l1 | LR_l2 | RF | SVM | GBDT | MLP |
|---|---|---|---|---|---|---|---|
| One year | 0.617 | 0.616 | 0.617 | 0.636 | 0.612 | 0.653 | 0.571 |
| Full history | 0.635 | 0.644 | 0.645 | 0.624 | 0.643 | 0.654 | 0.627 |




(a) Different time fusion methods.




(b) Different embedding strategies.

Xu H, et al. J Am Med Inform Assoc 2015;22:179–191. doi:10.1136/amiajnl-2014-002649, Research and Applications

# Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality

RECEIVED 15 January 2014
REVISED 10 June 2014
ACCEPTED 3 July 2014
PUBLISHED ONLINE FIRST 22 July 2014



Hua Xu[1], Melinda C Aldrich[2,3], Qingxia Chen[4,5], Hongfang Liu[6], Neeraja B Peterson[7], Qi Dai[3], Mia Levy[5,7], Anushi Shah[5], Xue Han[4], Xiaoyang Ruan[6], Min Jiang[1], Ying Li[8], Jamii St Julien[2], Jeremy Warner[5,7], Carol Friedman[8], Dan M Roden[7,9], Joshua C Denny[5,7]

**Figure 3**: Kaplan–Meier (K–M) plot of overall cancer survival for the Vanderbilt and Mayo Clinic cohorts. DM2, type 2 diabetes mellitus.

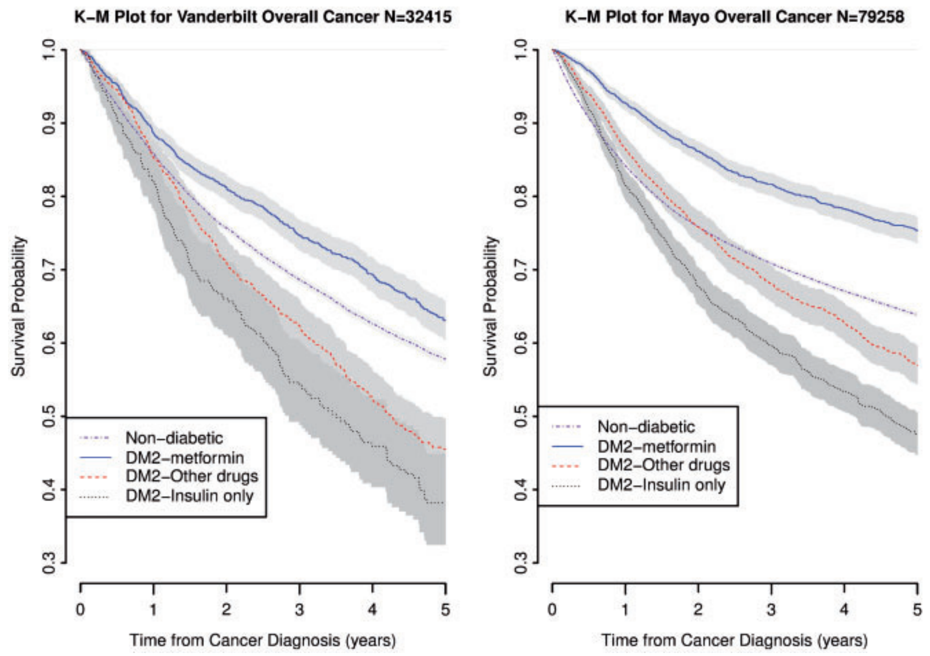



**Figure 1**: The study design and data extraction workflow for patients in the Vanderbilt electronic health record (EHR) system from January 1995 to December 2010.

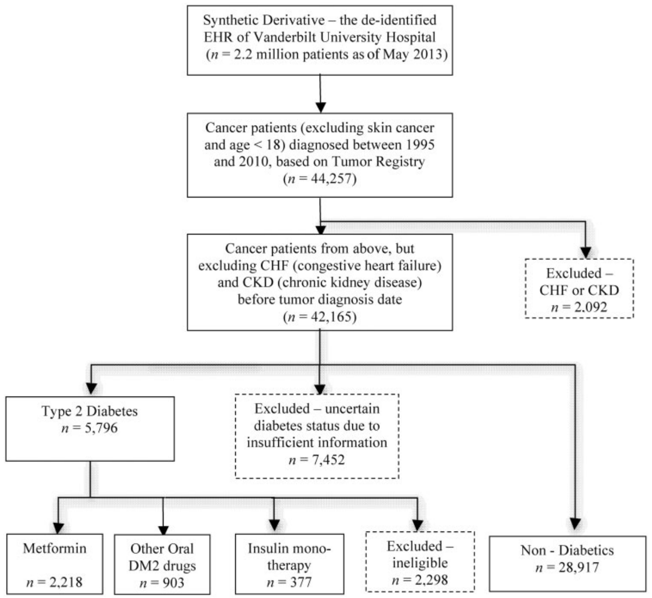
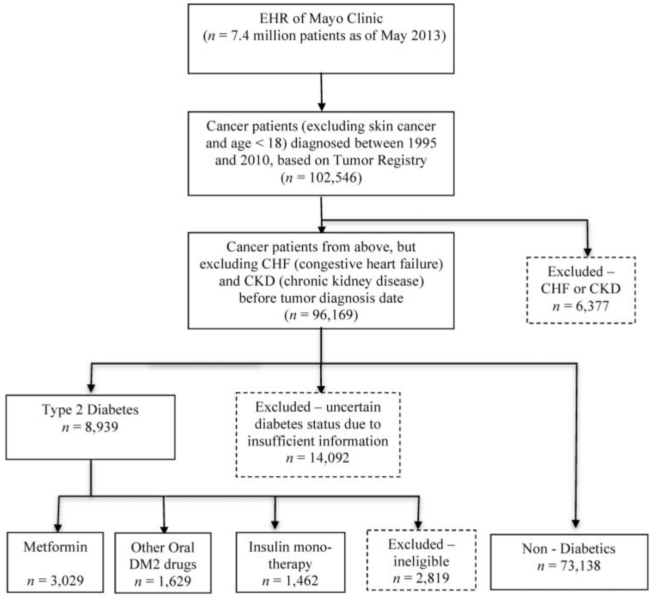



**Figure 2**: The study design and data extraction workflow for patients in the Mayo Clinic electronic health record (EHR) system from January 1995 to December 2010.



# Extended "Big Data" Analysis

- Single Drug


153 drugs for chronic diseases

- Carefully selected covariates (30+)

all-inclusive covariates (3000+)

- Hypothesis Driven

Data Driven

| Rank | Drug | HR | P-value | Lower | Upper | Adj-p-value |
|---|---|---|---|---|---|---|
| 1 | sildenafil | 0.704 | 0.000 | 0.621 | 0.797 | 0.000 |
| 2 | olmesartan | 0.714 | 0.001 | 0.580 | 0.879 | 0.017 |
| 3 | thyroxine | 0.791 | 0.000 | 0.734 | 0.852 | 0.000 |
| 4 | carvedilol | 0.803 | 0.003 | 0.693 | 0.930 | 0.033 |
| 5 | alendronic.acid | 0.812 | 0.000 | 0.725 | 0.908 | 0.004 |
| 6 | amlodipine | 0.839 | 0.000 | 0.782 | 0.901 | 0.000 |
| 7 | epoetin.alfa.recom | 0.840 | 0.000 | 0.772 | 0.914 | 0.001 |
| 8 | ramipril | 0.843 | 0.004 | 0.750 | 0.947 | 0.036 |
| 9 | simvastatin | 0.845 | 0.000 | 0.788 | 0.906 | 0.000 |
| 10 | atorvastatin | 0.861 | 0.000 | 0.797 | 0.930 | 0.003 |
| 11 | metformin | 0.862 | 0.003 | 0.781 | 0.952 | 0.033 |
| 12 | esomeprazole | 0.878 | 0.000 | 0.819 | 0.942 | 0.004 |
| 13 | omeprazole | 0.879 | 0.000 | 0.830 | 0.931 | 0.000 |
| 14 | lisinopril | 0.902 | 0.002 | 0.846 | 0.962 | 0.018 |
| 15 | lansoprazole | 0.908 | 0.005 | 0.849 | 0.971 | 0.037 |

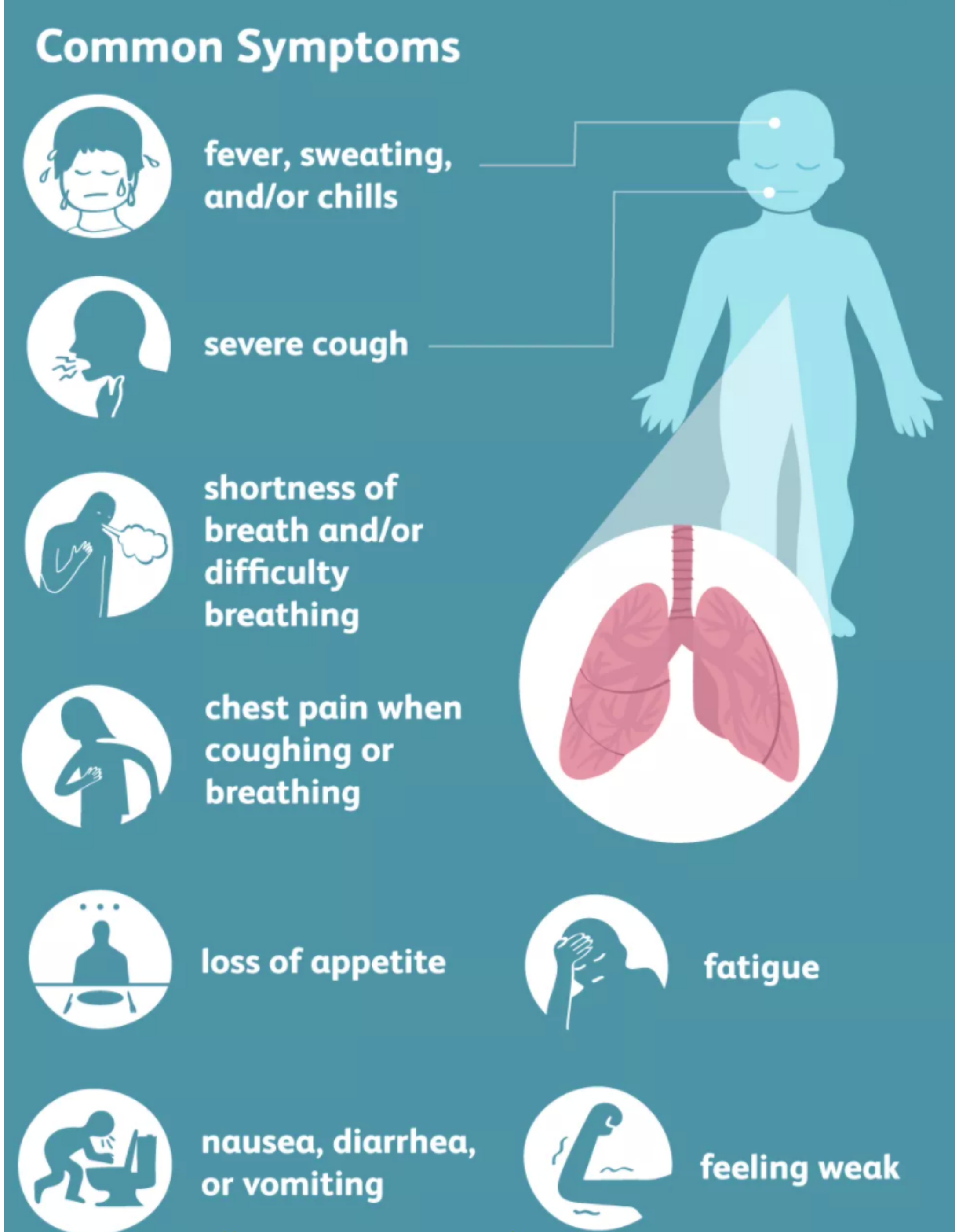
From Hua Xu's Slides

# Pneumonia & Asthma

A rule based machine learning algorithm suggested considering patients with both pneumonia and asthma to be at a lower risk of death from pneumonia than patients with pneumonia but with out asthma

At the hospitals hosting this study, patients with a history of asthma who presented with pneumonia were usually admitted directly to intensive care units to prevent complications, this led to patients with pneumonia and asthma having better outcomes than patients diagnosed with pneumonia and without the history of asthma, with an approximately 50% mortality risk reduction (5.4% vs. 11.3%)


Caruana, Rich, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noemie Elhadad. "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission." In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1721-1730. 2015.


https://www.verywellhealth.com/can-asthma-cause-pneumonia-200525

ORIGINAL ARTICLE

# Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19

The REMAP-CAP Investigators*


RESULTS

Both tocilizumab and sarilumab met the predefined criteria for efficacy. At that time, 353 patients had been assigned to tocilizumab, 48 to sarilumab, and 402 to control. The median number of organ support–free days was 10 (interquartile range, –1 to 16) in the tocilizumab group, 11 (interquartile range, 0 to 16) in the sarilumab group, and 0 (interquartile range, –1 to 15) in the control group. The median adjusted cumulative odds ratios were 1.64 (95% credible interval, 1.25 to 2.14) for tocilizumab and 1.76 (95% credible interval, 1.17 to 2.91) for sarilumab as compared with control, yielding posterior probabilities of superiority to control of more than 99.9% and of 99.5%, respectively. An analysis of 90-day survival showed improved survival in the pooled interleukin-6 receptor antagonist groups, yielding a hazard ratio for the comparison with the control group of 1.61 (95% credible interval, 1.25 to 2.08) and a posterior probability of superiority of more than 99.9%. All secondary analyses supported efficacy of these interleukin-6 receptor antagonists.

CONCLUSIONS

In critically ill patients with Covid-19 receiving organ support in ICUs, treatment with the interleukin-6 receptor antagonists tocilizumab and sarilumab improved outcomes, including survival. (REMAP-CAP ClinicalTrials.gov number, NCT02735707.)


The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

# Tocilizumab in Hospitalized Patients with Severe Covid-19 Pneumonia

I.O. Rosas, N. Bräu, M. Waters, R.C. Go, B.D. Hunter, S. Bhagani, D. Skiest, M.S. Aziz, N. Cooper, I.S. Douglas, S. Savic, T. Youngstein, L. Del Sorbo, A. Cubillo Gracian, D.J. De La Zerda, A. Ustianowski, M. Bao, S. Dimonaco, E. Graham, B. Matharu, H. Spotswood, L. Tsai, and A. Malhotra


RESULTS

Of the 452 patients who underwent randomization, 438 (294 in the tocilizumab group and 144 in the placebo group) were included in the primary and secondary analyses. The median value for clinical status on the ordinal scale at day 28 was 1.0 (95% confidence interval [CI], 1.0 to 1.0) in the tocilizumab group and 2.0 (non-ICU hospitalization without supplemental oxygen) (95% CI, 1.0 to 4.0) in the placebo group (between-group difference, –1.0; 95% CI, –2.5 to 0; P=0.31 by the van Elteren test). In the safety population, serious adverse events occurred in 103 of 295 patients (34.9%) in the tocilizumab group and in 55 of 143 patients (38.5%) in the placebo group. Mortality at day 28 was 19.7% in the tocilizumab group and 19.4% in the placebo group (weighted difference, 0.3 percentage points (95% CI, –7.6 to 8.2; nominal P=0.94).

CONCLUSIONS

In this randomized trial involving hospitalized patients with severe Covid-19 pneumonia, the use of tocilizumab did not result in significantly better clinical status or lower mortality than placebo at 28 days. (Funded by F. Hoffmann–La Roche and the Department of Health and Human Services; COVACTA ClinicalTrials.gov number, NCT04320615.)

# The Future

# Understand the Data

thebmj Research Education News & Views Campaigns Jobs

**Research**

## Biases in electronic health record data due to processes within the healthcare system: retrospective observational study

*BMJ* 2018 ; 361 doi: https://doi.org/10.1136/bmj.k1479 (Published 30 April 2018)
Cite this as: *BMJ* 2018;361:k1479

*Denis Agniel, research fellow* [1], *Isaac S Kohane, department head* [1,2], *Griffin M Weber, associate professor* [1,3]

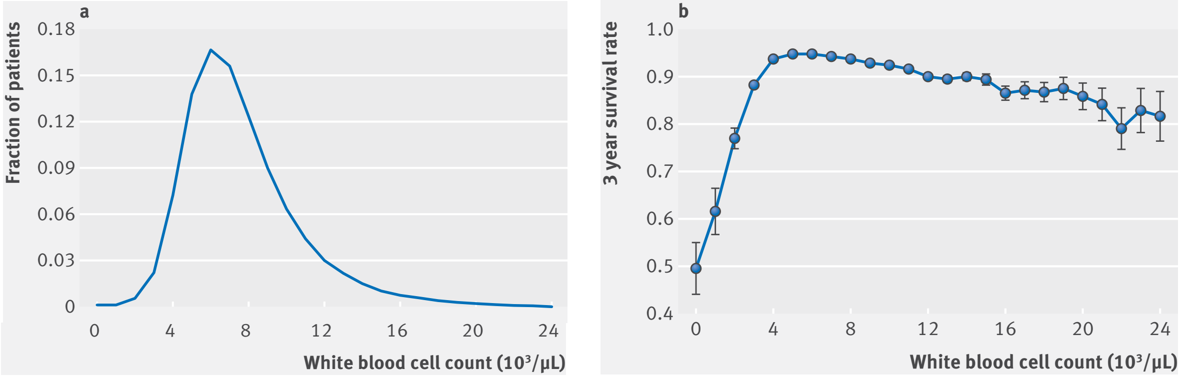
Author affiliations

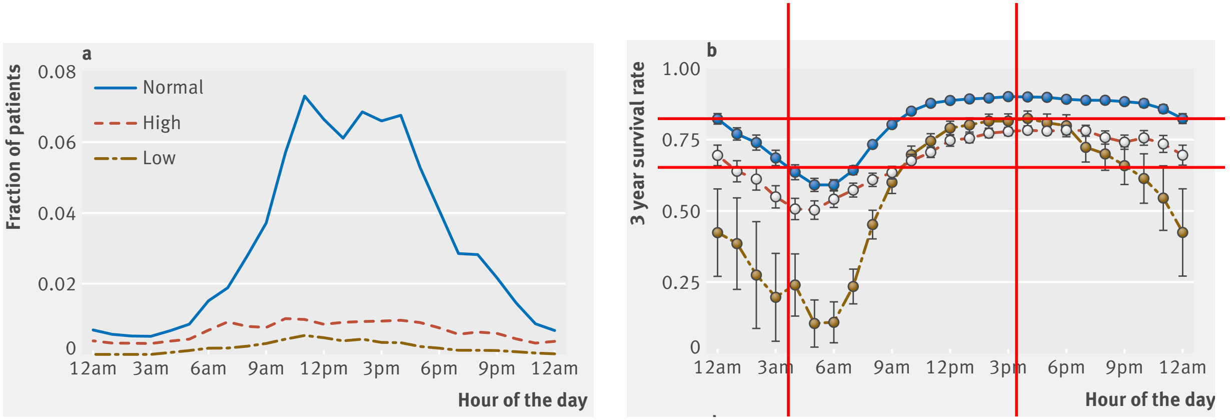
[1] *Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St, Boston, MA 02115, USA*

[2] *Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA*

[3] *Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA*

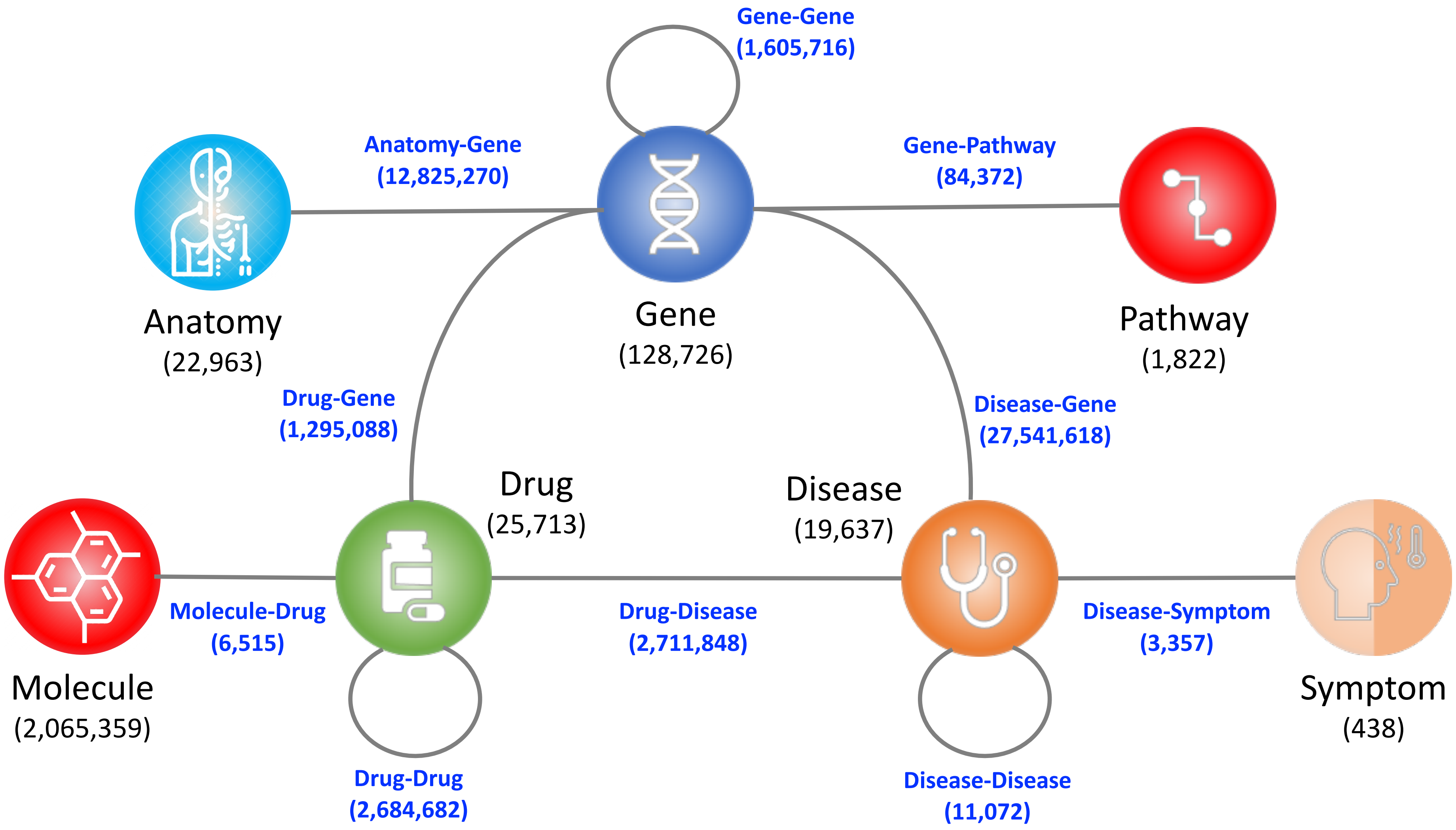
Correspondence to: G M Weber weber@hms.harvard.edu




https://www.bmj.com/content/361/bmj.k1479

# Knowledge is Important



# CLASSIFYING CLINICALLY ACTIONABLE GENETIC MUTATIONS

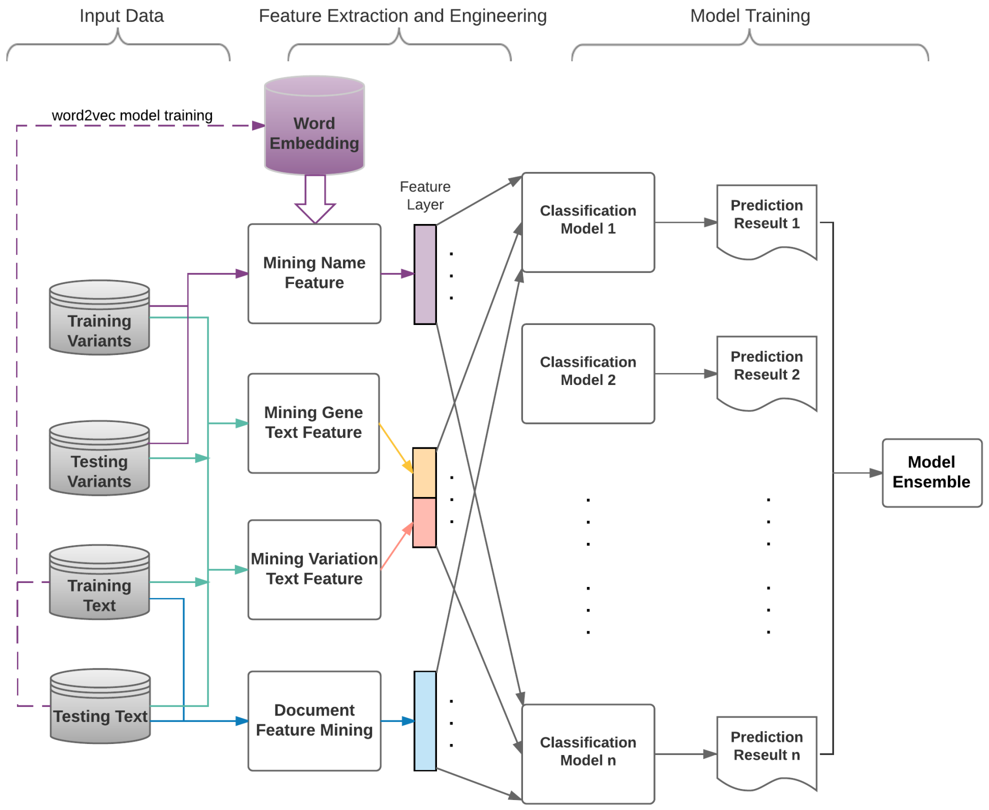
## NIPS 2017 COMPETITION

Once sequenced, a cancer tumor can have thousands of ***genetic mutations***. But the challenge is distinguishing the mutations that contribute to tumor growth (drivers) from the neutral mutations (passengers).

Currently this interpretation of genetic mutations is being done ***manually***. This is a very time-consuming task where a clinical pathologist has to manually review and classify every single genetic mutation based on evidence from ***text-based clinical literature***.

For this competition MSKCC is making available an expert-annotated knowledge base where world-class researchers and oncologists have ***manually*** annotated thousands of mutations.

The goal of this competition is to develop a Machine Learning algorithm that, using this knowledge base as a baseline, ***automatically*** classifies genetic variations.




https://www.mskcc.org/news/msk-advances-its-ai-machine-learning-nips-2017


Zhang, Xi, Dandi Chen, Yongjun Zhu, Chao Che, Chang Su, Sendong Zhao, Xu Min, and Fei Wang. "Multi-view ensemble classification for clinically actionable genetic mutations." In *The NIPS'17 Competition: Building Intelligent Systems*, pp. 79-99. Springer, Cham, 2018.

# Model Underspecification

arXiv:2011.03395v2 [cs.LG] 24 Nov 2020

Underspecification Presents Challenges for Credibility in Modern Machine Learning


| | |
|---|---|
| Alexander D'Amour* | ALEXDAMOUR@GOOGLE.COM |
| Katherine Heller* | KHELLER@GOOGLE.COM |
| Dan Moldovan* | MDAN@GOOGLE.COM |
| Ben Adlam | ADLAM@GOOGLE.COM |
| Babak Alipanahi | BABAKA@GOOGLE.COM |
| Alex Beutel | ALEXBEUTEL@GOOGLE.COM |
| Christina Chen | CHRISTINIUM@GOOGLE.COM |
| Jonathan Deaton | JDEATON@GOOGLE.COM |
| Jacob Eisenstein | JEISENSTEIN@GOOGLE.COM |
| Matthew D. Hoffman | MHOFFMAN@GOOGLE.COM |
| Farhad Hormozdiari | FHORMOZ@GOOGLE.COM |
| Neil Houlsby | NEILHOULSBY@GOOGLE.COM |
| Shaobo Hou | SHAOBOHOU@GOOGLE.COM |
| Ghassen Jerfel | GHASSEN@GOOGLE.COM |
| Alan Karthikesalingam | ALANKARTHI@GOOGLE.COM |
| Mario Lucic | LUCIC@GOOGLE.COM |
| Yian Ma | YIANMA@UCSD.EDU |
| Cory McLean | CYM@GOOGLE.COM |
| Diana Mincu | DMINCU@GOOGLE.COM |
| Akinori Mitani | AMITANI@GOOGLE.COM |
| Andrea Montanari | MONTANARI@STANFORD.EDU |
| Zachary Nado | ZNADO@GOOGLE.COM |
| Vivek Natarajan | NATVIV@GOOGLE.COM |
| Christopher Nielson† | CHRISTOPHER.NIELSON@VA.GOV |
| Thomas F. Osborne† | THOMAS.OSBORNE@VA.GOV |
| Rajiv Raman | DRRRN@SNMAIL.ORG |
| Kim Ramasamy | KIM@ARAVIND.ORG |
| Rory Sayres | SAYRES@GOOGLE.COM |
| Jessica Schrouff | SCHROUFF@GOOGLE.COM |
| Martin Seneviratne | MARTSEN@GOOGLE.COM |
| Shannon Sequeira | SHNNN@GOOGLE.COM |
| Harini Suresh | HSURESH@MIT.EDU |
| Victor Veitch | VICTORVEITCH@GOOGLE.COM |
| Max Vladymyrov | MXV@GOOGLE.COM |
| Xuezhi Wang | XUEZHIW@GOOGLE.COM |
| Kellie Webster | WEBSTERK@GOOGLE.COM |
| Steve Yadlowsky | YADLOWSKY@GOOGLE.COM |
| Taedong Yun | TEDYUN@GOOGLE.COM |
| Xiaohua Zhai | XZHAI@GOOGLE.COM |
| D. Sculley | DSCULLEY@GOOGLE.COM |


the solution to a problem is underspecified if there are many distinct solutions that solve the problem equivalently

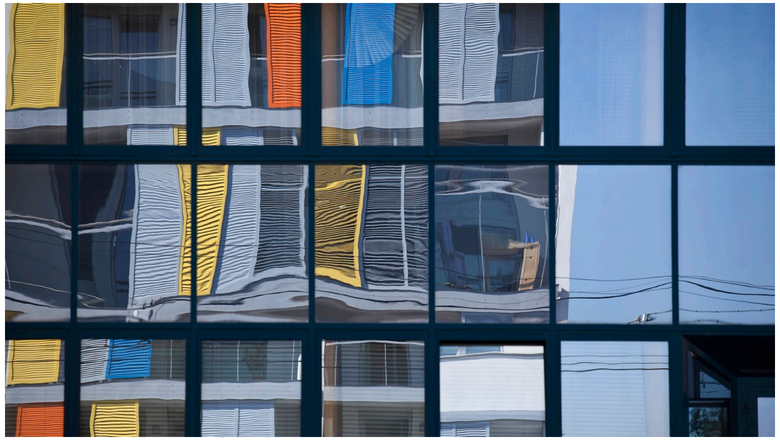
MIT Technology Review

Artificial intelligence / Machine learning

## The way we train AI is fundamentally flawed

The process used to build most of the machine-learning models we use today can't tell if they will work in the real world or not—and that's a problem.

by **Will Douglas Heaven** November 18, 2020



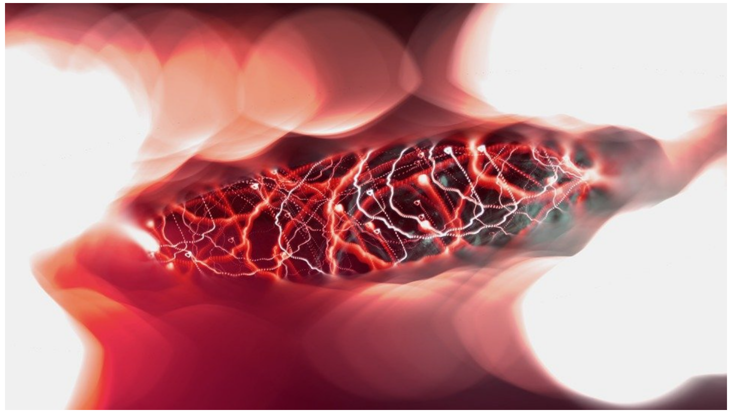
https://www.technologyreview.com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision/

UNITE.AI

ARTIFICIAL NEURAL NETWORKS

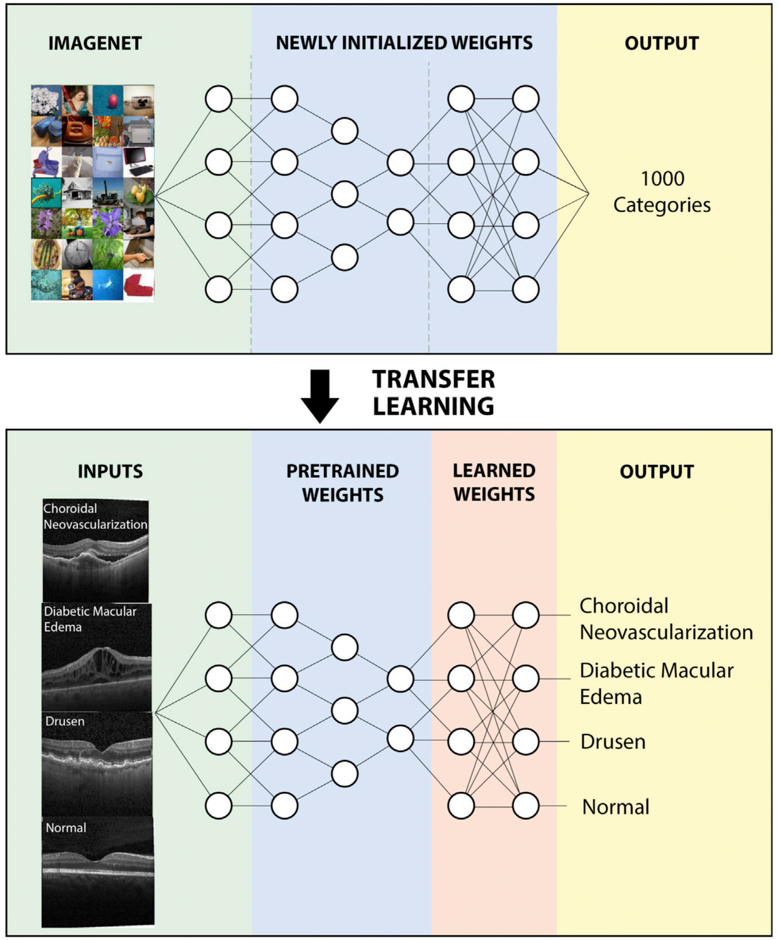
## Google Researchers Discover Underspecification Problem Holding Bacl Many AI Models

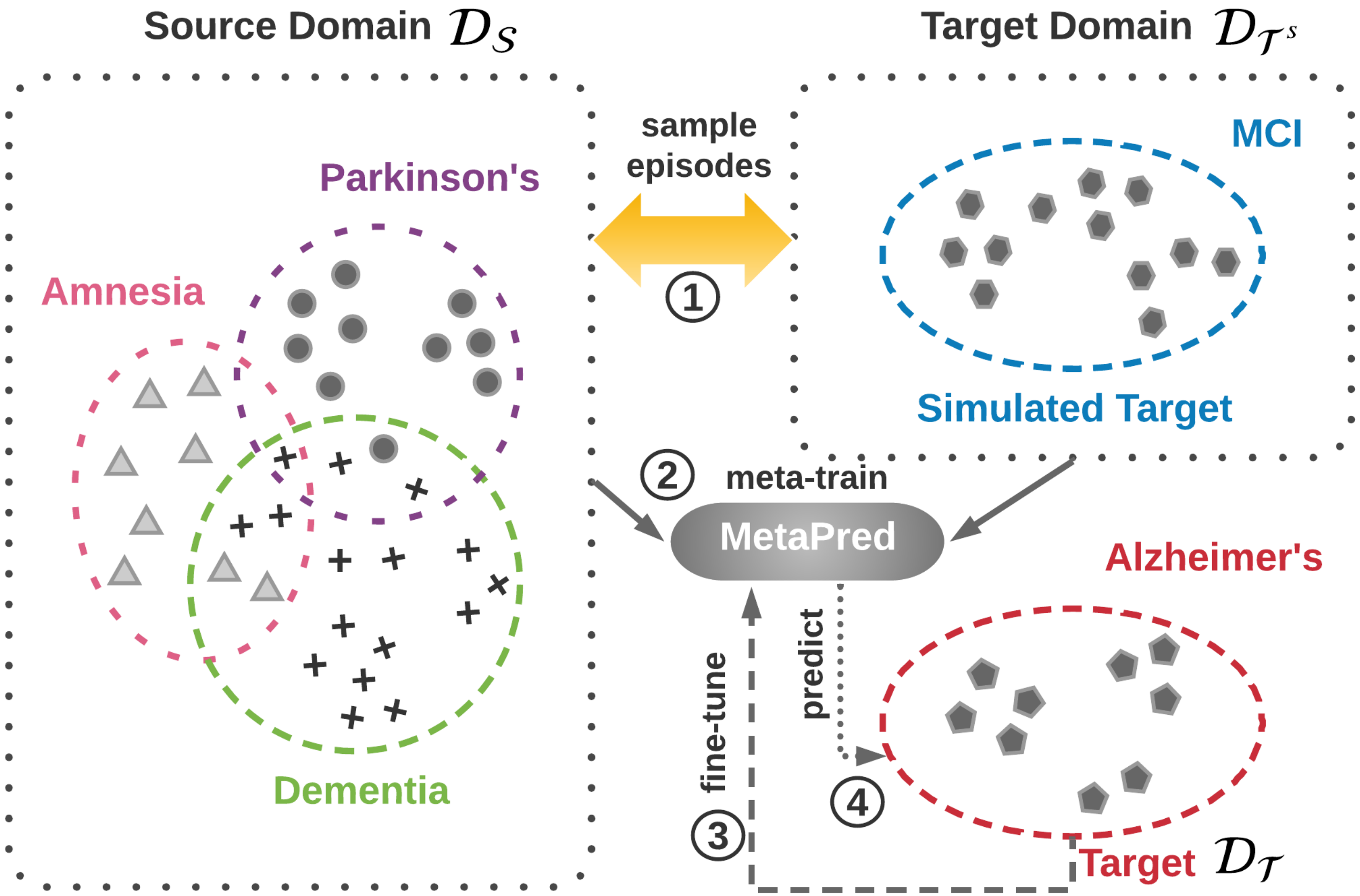
Published 1 week ago on November 20, 2020
By Daniel Nelson



https://www.unite.ai/google-researchers-discover-underspecification-problem-holding-back-many-ai-models/

# Transfer Learning

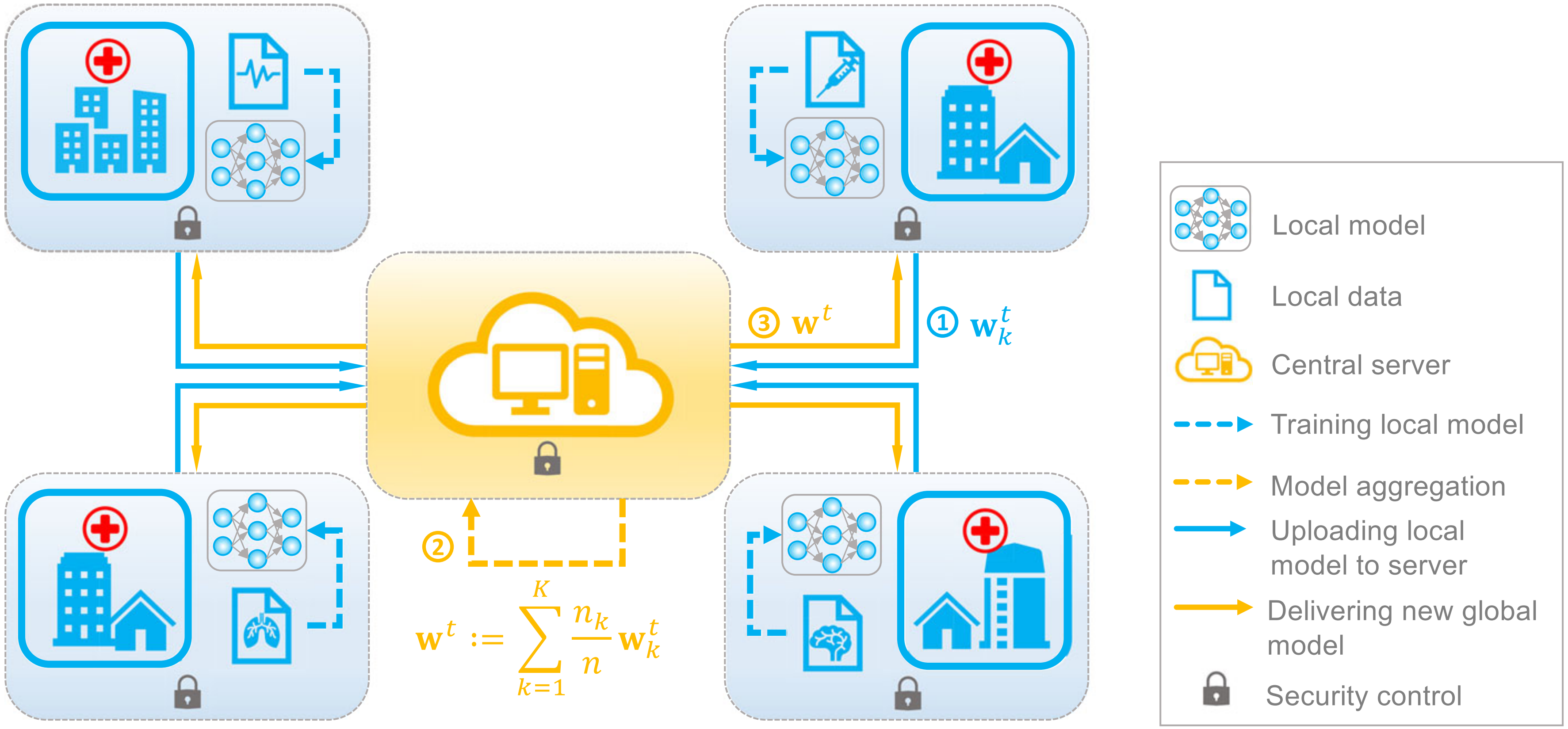



Kermany, Daniel S., Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning." *Cell* 172, no. 5 (2018): 1122-1131.

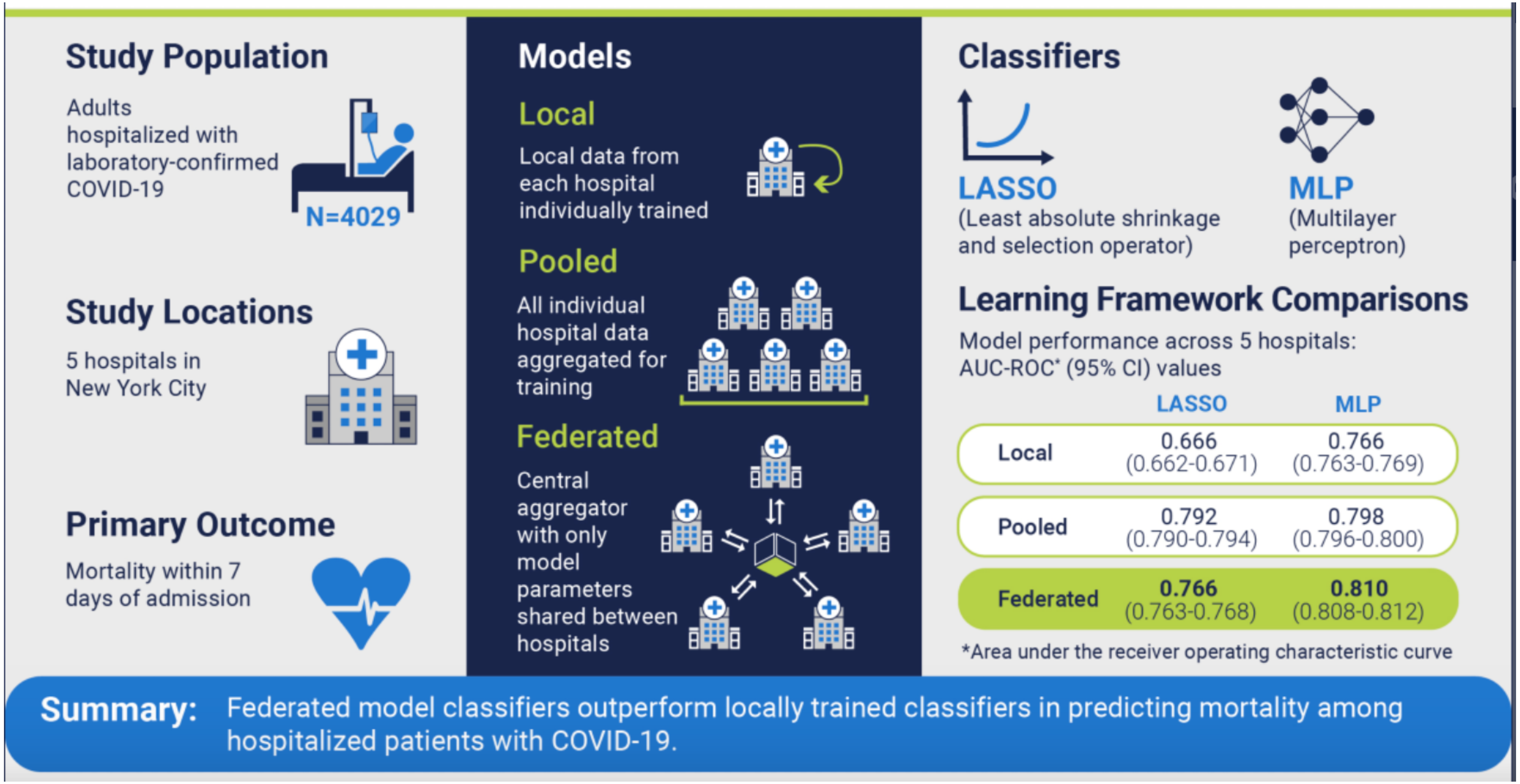
Zhang, Xi Sheryl, Fengyi Tang, Hiroko H. Dodge, Jiayu Zhou, and ***Fei Wang***. "Metapred: Meta-learning for clinical risk prediction with limited patient electronic health records." In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2487-2495. 2019.

# Federated Learning



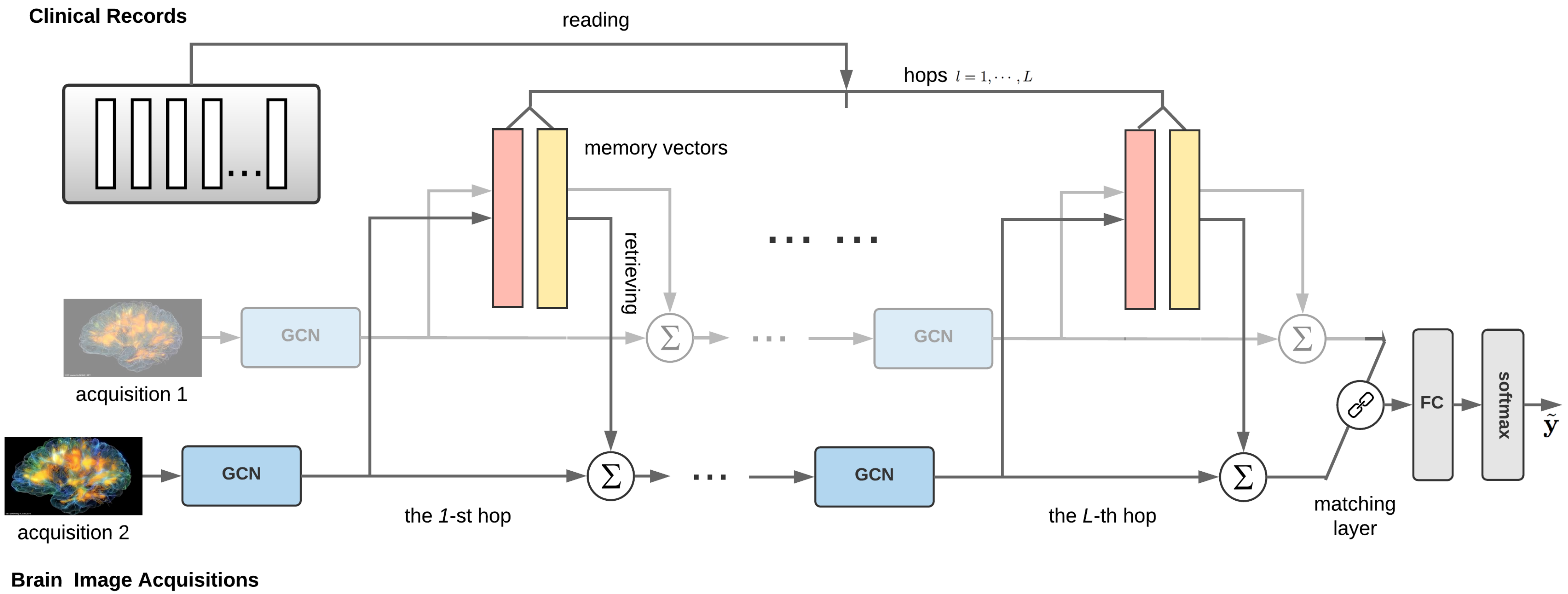

Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and ***Fei Wang***. "Federated learning for healthcare informatics." *Journal of Healthcare Informatics Research* (2020): 1-19.

# Federated Learning



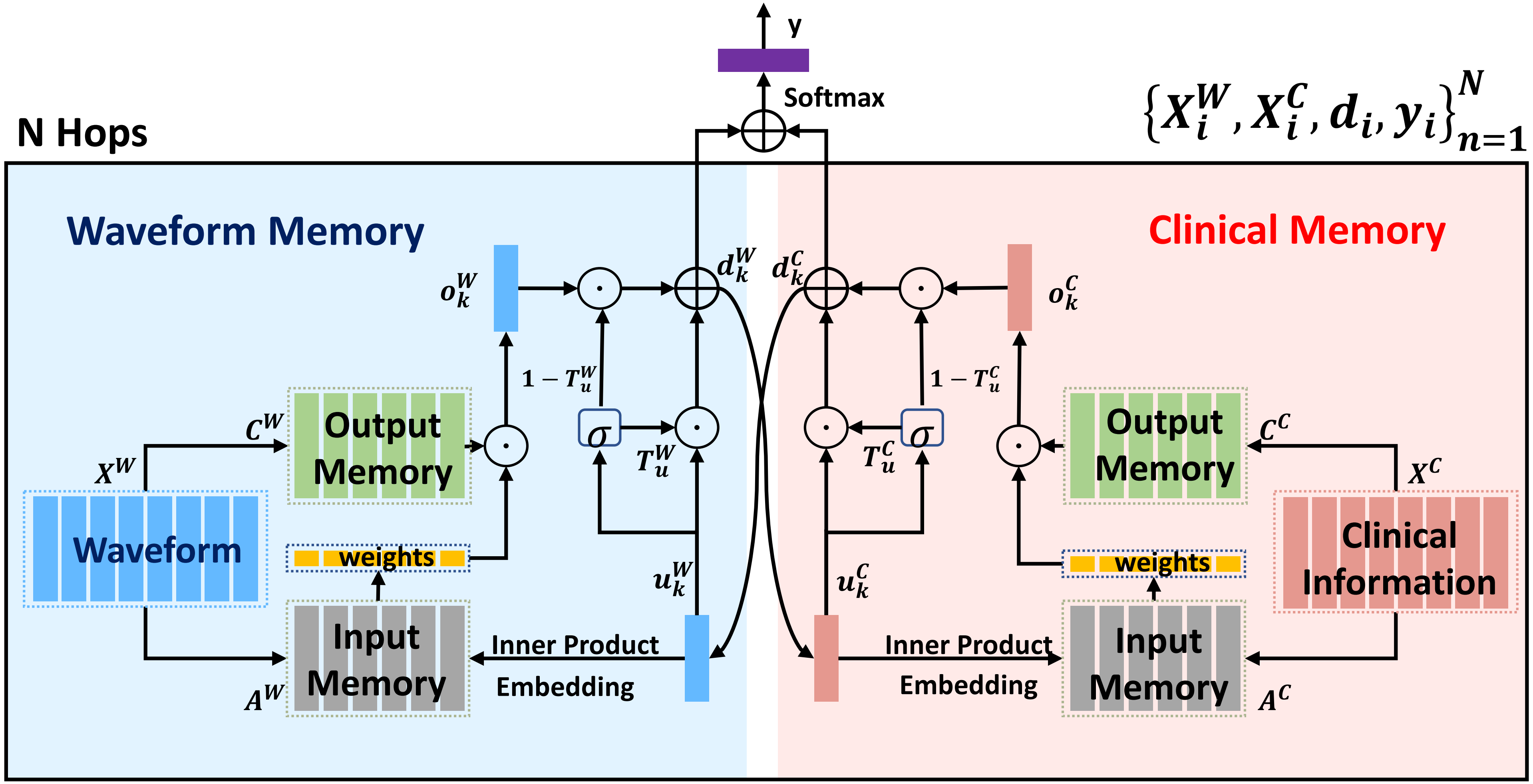

Vaid, Akhil, Suraj K. Jaladanki, Jie Xu, Shelly Teng, Arvind Kumar, Samuel Lee, Sulaiman Somani, Jessica K De Freitas, Tingyi Wanyan, Kipp W Johnson, Mesude Bicak, Eyal Klang, Young Joon Kwon, Anthony Costa, Shan Zhao, Riccardo Miotto, Alexander W Charney, Erwin Böttinger, Zahi A Fayad, Girish N Nadkarni, ***Fei Wang***, Benjamin S Glicksberg. “Federated learning of electronic health records improves mortality prediction in patients hospitalized with covid-19.” *JMIR Medical Informatics* (2021).

# Multi-Modal Learning




Zhang, Xi, Jingyuan Chou, and ***Fei Wang***. "Integrative analysis of patient health records and neuroimages via memory-based graph convolutional network." In 2018 IEEE International Conference on Data Mining (ICDM), pp. 767-776. IEEE, 2018.

# Multi-Modal Learning





Feng, Yujuan, Zhenxing Xu, Lin Gan, Ning Chen, Bin Yu, Ting Chen, and ***Fei Wang***. "DCMN: Double Core Memory Network for Patient Outcome Prediction with Multimodal Data." In *2019 IEEE International Conference on Data Mining (ICDM)*, pp. 200-209. IEEE, 2019.

# Model Interpretation

**Annals of Internal Medicine®**

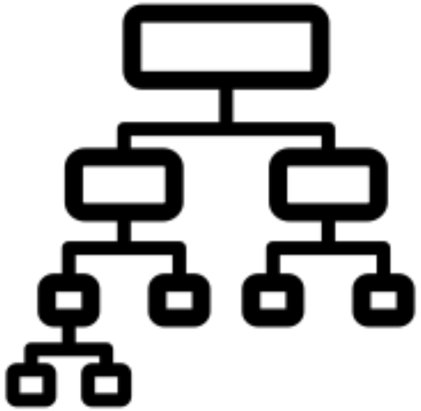
LATEST ISSUES CHANNELS CME/MOC IN THE CLINIC JOURNAL CLUB WEB EXCLUSIVES AUTHOR INFO

‹ PREV ARTICLE | THIS ISSUE | NEXT ARTICLE ›

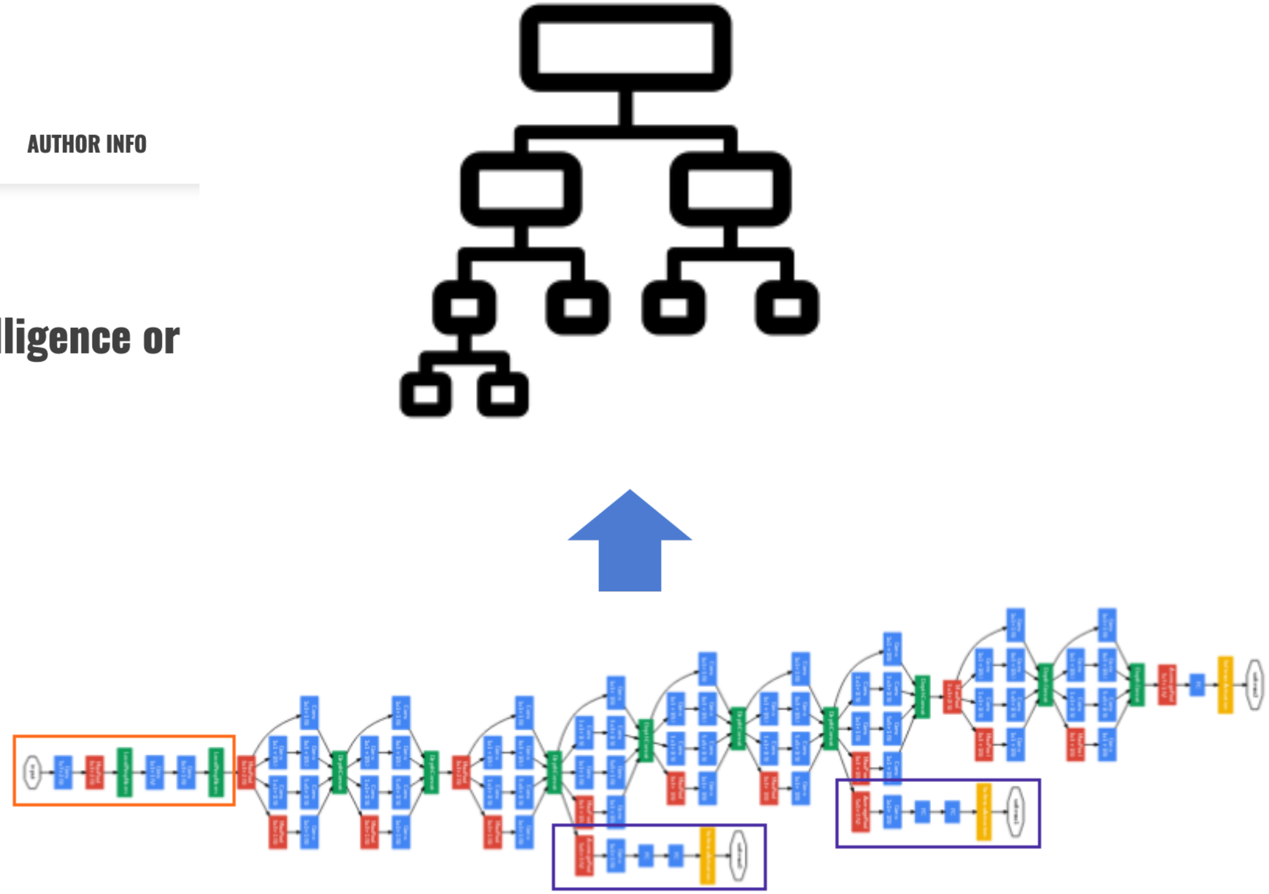
IDEAS AND OPINIONS | 7 JANUARY 2020

## Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine?

*Fei Wang, PhD; Rainu Kaushal, MD, MPH; Dhruv Khullar, MD, MPP*



**Intrinsic Interpretability**



**Post-hoc Interpretability**

# Attention Is All You Need

**Ashish Vaswani*** Google Brain avaswani@google.com
**Noam Shazeer*** Google Brain noam@google.com
**Niki Parmar*** Google Research nikip@google.com
**Jakob Uszkoreit*** Google Research usz@google.com
**Llion Jones*** Google Research llion@google.com
**Aidan N. Gomez*** † University of Toronto aidan@cs.toronto.edu
**Łukasz Kaiser*** Google Brain lukaszkaiser@google.com
**Illia Polosukhin*** ‡ illia.polosukhin@gmail.com

## Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

# Attention is not Explanation

**Sarthak Jain** Northeastern University jain.sar@husky.neu.edu
**Byron C. Wallace** Northeastern University b.wallace@northeastern.edu

# Attention is not not Explanation

**Sarah Wiegreffe*** School of Interactive Computing Georgia Institute of Technology saw@gatech.edu
**Yuval Pinter*** School of Interactive Computing Georgia Institute of Technology uvp@gatech.edu

# Why is Attention Not So Interpretable?

**Bing Bai[1*], Jian Liang[2*], Guanhua Zhang[13], Hao Li[1], Kun Bai[1], Fei Wang[4]**
[1]Tencent Inc., China, [2]Alibaba Inc., China,
[3]Harbin Institute of Technology, China, [4]Cornell University, USA
{icebai,guanhzhang,leehaoli,kunbai}@tencent.com,
xuelang.lj@alibaba-inc.com, few2001@med.cornell.edu

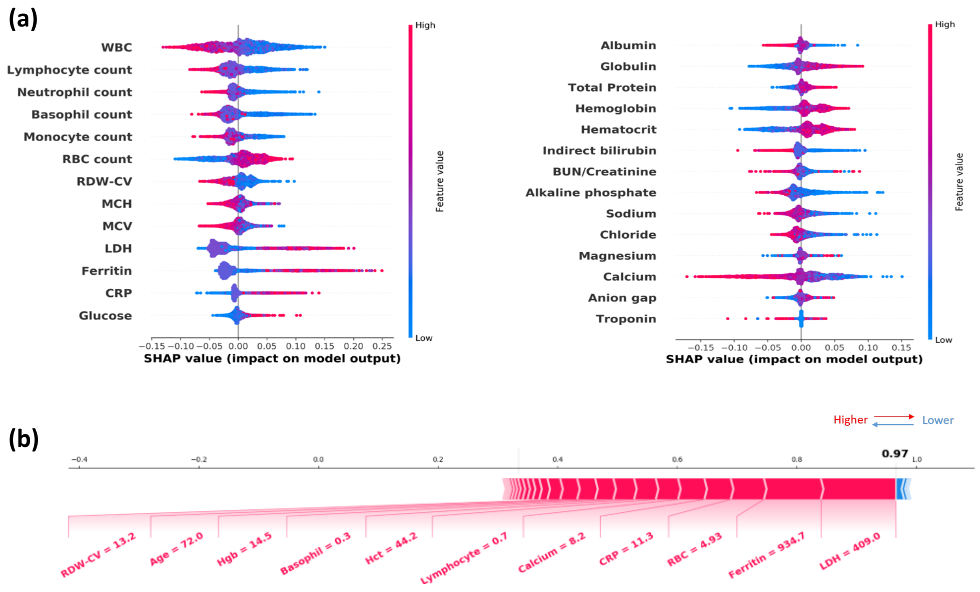
OXFORD
ACADEMIC

Clinical Chemistry

# Routine Laboratory Blood Tests Predict SARS-CoV-2 Infection Using Machine Learning FREE

He S Yang, Yu Hou, Ljiljana V Vasovic, Peter A D Steel, Amy Chadburn, Sabrina E Racine-Brzostek, Priya Velu, Melissa M Cushing, Massimo Loda, Rainu Kaushal
... Show more
Author Notes

*Clinical Chemistry*, Volume 66, Issue 11, November 2020, Pages 1396–1404,
https://doi.org/10.1093/clinchem/hvaa200
**Published:** 30 October 2020 **Article history ▾**




# Approximation Trees: Statistical Stability in Model Distillation

Yichen Zhou, Zhengze Zhou, Giles Hooker
Department of Statistical Science
Cornell University
Ithaca, NY 14853, USA

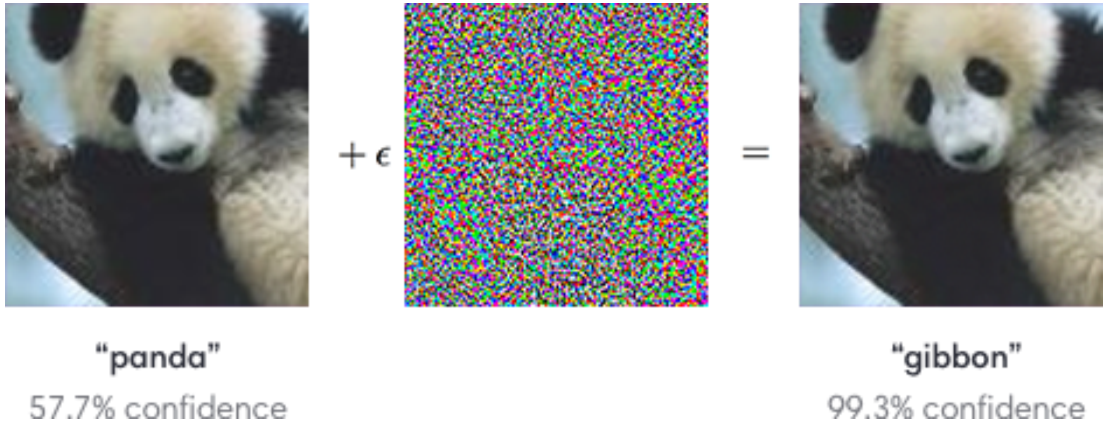
**Abstract**

This paper examines the stability of learned explanations for black-box predictions via model distillation with decision trees. One approach to intelligibility in machine learning is to use an understandable "student" model to mimic the output of an accurate "teacher". Here, we consider the use of regression trees as a student model, in which nodes of the tree can be used as "explanations" for particular predictions, and the whole structure of the tree can be used as a global representation of the resulting function.

However, individual trees are sensitive to the particular data sets used to train them, and an interpretation of a student model may be suspect if small changes in the training data have a large effect on it. In this context, access to outcomes from a teacher helps to stabilize the greedy splitting strategy by generating a much larger corpus of training examples than was originally available. We develop tests to ensure that enough examples are generated at each split so that the same splitting rule would be chosen with high probability were the tree to be re-trained. Further, we develop a stopping rule to indicate how deep the tree should be built based on recent results on the variability of Random Forests when these are used as the teacher. We provide concrete examples of these procedures on the CAD-MDD and COMPAS data sets.

# Model Security

**Adversarial machine learning** is a technique employed in the field of machine learning which attempts to ***fool models through malicious input***. This technique can be applied for a variety of reasons, the most common being to attack or cause a malfunction in standard machine learning models.




https://arxiv.org/abs/1412.6572


CAR AND DRIVER REVIEWS NEWS FEATURES BUYER'S GUIDE SHOP FOR CARS SUBSCRIBE NEWSLETTER

**Researchers Find a Malicious Way to Meddle with Autonomous Cars**

MARK HARRIS AUG 4, 2017



https://www.caranddriver.com/news/a15340148/researchers-find-a-malicious-way-to-meddle-with-autonomous-cars/

Science Contents News Careers Journals

SHARE

POLICY FORUM | MACHINE LEARNING

# Adversarial attacks on medical machine learning

Samuel G. Finlayson[1], John D. Bowers[2], Joichi Ito[3], Jonathan L. Zittrain[2], Andrew L. Beam[4], Isaac S. Kohane[1]
+ See all authors and affiliations

*Science* 22 Mar 2019:
Vol. 363, Issue 6433, pp. 1287-1289
DOI: 10.1126/science.aaw4399



## The anatomy of an adversarial attack

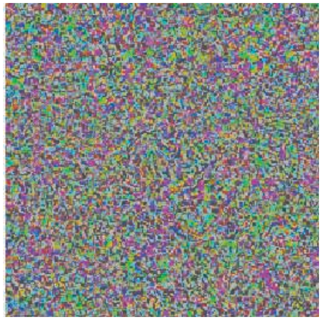
Demonstration of how adversarial attacks against various medical AI systems might be executed without requiring any overtly fraudulent misrepresentation of the data.

**Original image**

Dermatoscopic image of a benign melanocytic nevus, along with the diagnostic probability computed by a deep neural network.

**+ 0.04 ×**

**Adversarial noise**



Perturbation computed by a common adversarial attack technique. See (*7*) for details.

**=**

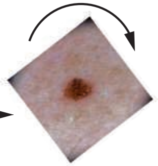
**Adversarial example**

Combined image of nevus and attack perturbation and the diagnostic probabilities from the same deep neural network.

Model confidence

**Diagnosis: Benign**

**Adversarial rotation** (*8*) →

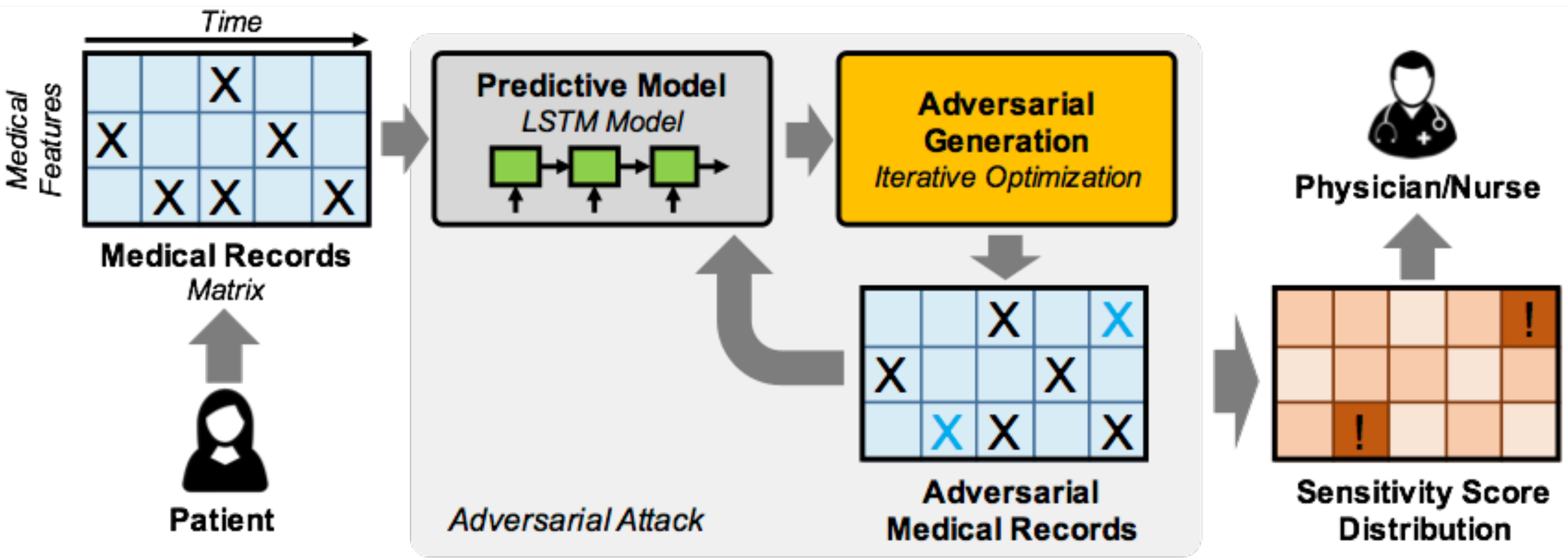


Model confidence

**Diagnosis: Malignant**

| | | |
| --- | --- | --- |
| The patient has a history of back pain and chronic alcohol abuse and more recently has been seen in several... | **Adversarial text substitution** (*9*) → | The patient has a history of lumbago and chronic alcohol dependence and more recently has been seen in several... |
| **Opioid abuse risk: High** | | **Opioid abuse risk: Low** |
| 277.7 Metabolic syndrome<br>429.9 Heart disease, unspecified<br>278.00 Obesity, unspecified | **Adversarial coding** (*13*) → | 401.0 Benign essential hypertension<br>272.0 Hypercholesterolemia<br>272.2 Hyperglyceridemia<br>429.9 Heart disease, unspecified<br>278.00 Obesity, unspecified |
| **Reimbursement: Denied** | | **Reimbursement: Approved** |

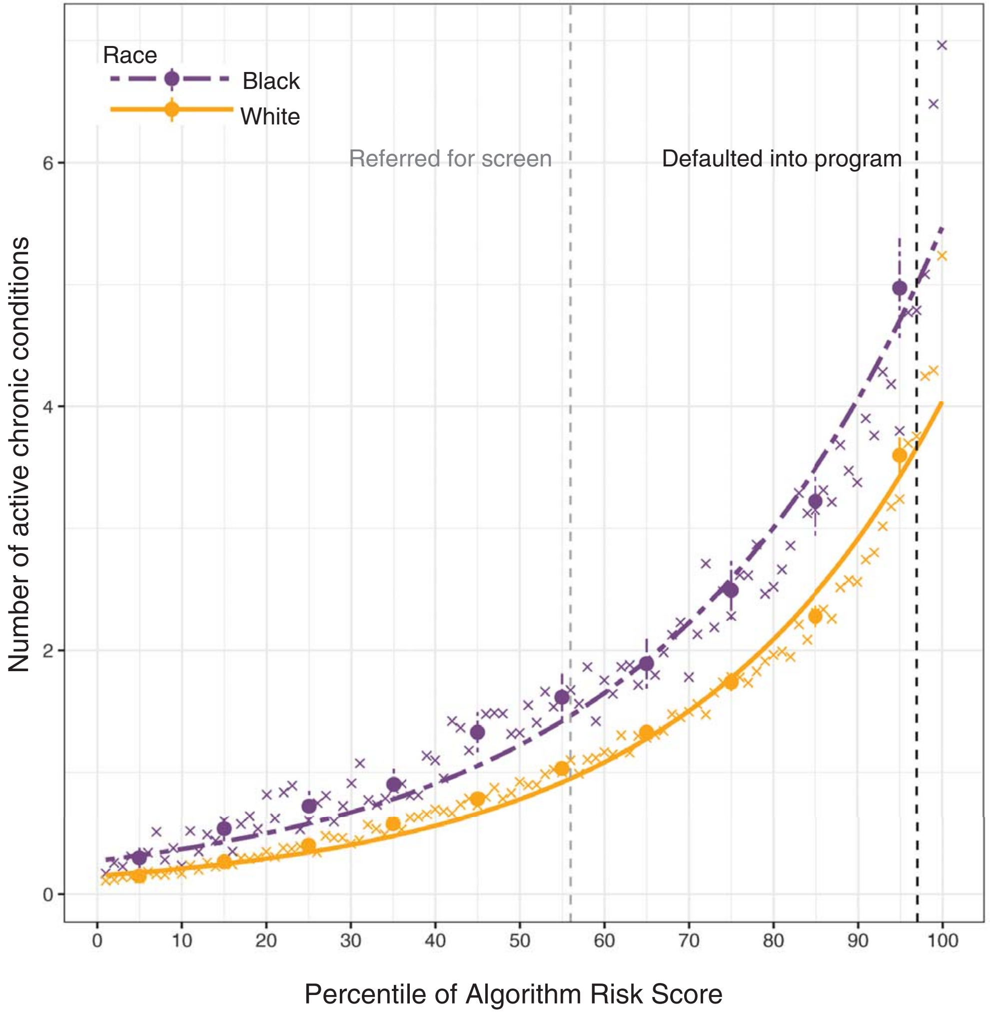
# Model Security




$$\min_{\tilde{X}} \max \left\{ \left[Logit(\tilde{X})\right]_{y_\theta} - \left[Logit(\tilde{X})\right]_{\tilde{y}_\theta}, -\kappa \right\} + \lambda \|\tilde{X} - X\|_1$$

Sun, Mengying, Fengyi Tang, Jinfeng Yi, ***Fei Wang***, and Jiayu Zhou. "Identify Susceptible Locations in Medical Records via Adversarial Attacks on Deep Predictive Models." *Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* ***(KDD)*** (2018).

# Model Bias




Science Contents News Careers Journals

SHARE

RESEARCH ARTICLE

## Dissecting racial bias in an algorithm used to manage the health of populations


Ziad Obermeyer[1,2,*], Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5,*,†]

+ See all authors and affiliations



*Science* 25 Oct 2019:
Vol. 366, Issue 6464, pp. 447-453
DOI: 10.1126/science.aax2342


## `xOrder`: A Model Agnostic Post-Processing Framework for Achieving Ranking Fairness While Maintaining Algorithm Utility

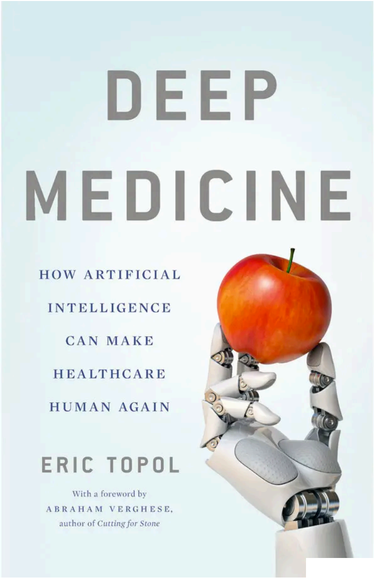

Sen Cui[1*] Weishen Pan[1*] Changshui Zhang[1] Fei Wang[2]

# Acknowledgement

- Current Members
  - Chang Su
  - Jie Xu
  - Zhenxing Xu
  - Chengxi Zang
  - Yu Hou
  - Hao Zhang
  - Mattew Brendel
- Alumni
  - Xi Zhang
  - Yongjun Zhu
  - Lifang He
  - Sendong Zhao
  - Jingyuan Chou
  - Dandi Chen
- Collaborators
  - Robert Aseltine (Psychiatry)
  - Daniel Barone (Neurology)
  - Larry Casalino (Health Policy)
  - Kun Chen (Statistics)
  - Hiroko Dodge (Neurology)
  - Olivier Elemento (System Biology)
  - James Flory (Endocrinology)
  - Parag Goyal (Cardiology)
  - Zachary Grinspan (Neurology)
  - Claire Henchcliffe (Neurology)
  - Giles Hooker (Statistics)
  - Richard Isaacson (Neurology)
  - Dhruv Khullar (Internal Medicine)
  - Anais Rameau (Otolaryngology)
  - Harini Sarva (Neurology)
  - Edward Schenck (Intensivist)
  - Sarina He Yang (Lab Medicine)

# Acknowledgement

NSF
NATIONAL INSTITUTES OF HEALTH
NAVAL MEDICAL RESEARCH CENTER
U.S. NAVY
Office of Naval Research
USN
ONR
Science & Technology
mitre
Boehringer Ingelheim
THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH
pcori®
American Heart Association®
Google
amazon web services
FORCECLOUDS
Air Liquide
OCEAN EX




"The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust—the human touch—between patients and doctors"


feiwang.cornell@gmail.com

feiwang03