



# Inaugural Artificial Intelligence and Data Science Year in Review

**Yuan Luo, PhD, FAMIA**

Northwestern University Feinberg School of Medicine

@yuanhypnosluo

**Hongfang Liu, PhD, FACMI**

Mayo Clinic

@hongfangliu

#IS23

# Disclosure

---



Industry: Walmart (Luo)

Funding Support: NIH

The content is solely our responsibility and does not necessarily represent the official views of our institutions or the National Institutes of Health.

Copyrights are reserved by the authors who provided the slides that we adapted from.

# Learning Objectives

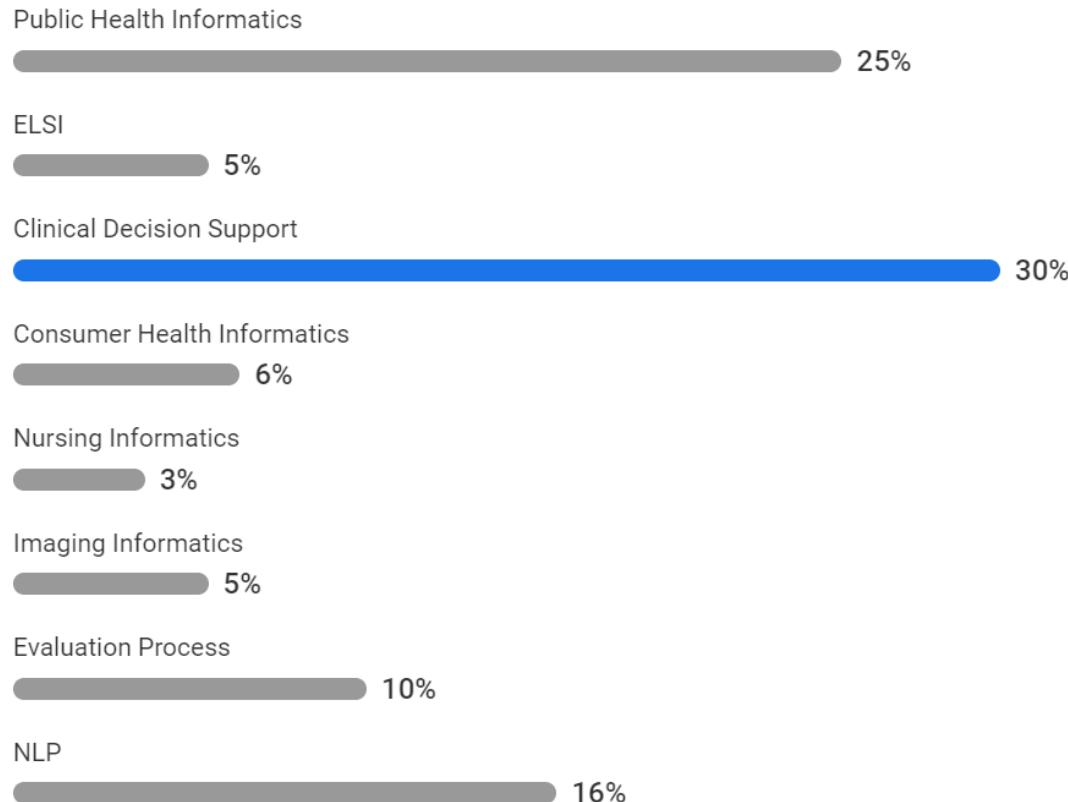
---



After participating in this session, the learner should be better able to:

- Have a sense of the scope of AI and Data Science for health care
- Summarize the state-of-the-art research in AI and Data Science for health care
- Identify research and development opportunities in this space and advance the field

# What best describes your interest (select one)?



# Process and caveats

---



The content is developed by the informatics community, for the informatics community, we are not boiling the ocean

Teamwork makes dreamwork, we thank talented volunteers from JAMIA and JBI student editorial boards++ (Brian Tran, Hong Sun, Xinmeng Zhang, Joshua Anderson, Yikuan Li, Elaine Sang, Hanyin Wang, Meghan Hutch), generous help from Jim Cimino

Each volunteer picked a topic, for which they were asked to nominate papers that are significant (new model, new insights, new task)

Focused on original research of AI and data science for health care, exclude reviews, viewpoints ...

I picked “interesting” ones to request audiovisual materials from authors

Hongfang and I finalized the selection, organized the slides and divide-and-conquer the presentation

Apologies for things I missed or misjudged, mistakes are all mine

# Topics

---



- Public and Global Health Informatics
- Ethical, Legal and Social Issues
- Clinical Decision Support
- Consumer and Pervasive Health Informatics
- Nursing Informatics
- Imaging Informatics
- Natural Language Processing
- Evaluation and Implementation

# Public and Global Health Informatics

› J Biomed Inform. 2022 May;129:104051. doi: 10.1016/j.jbi.2022.104051. Epub 2022 Mar 19.

## Supporting rural public health practice to address local-level social determinants of health across Northwest states: Development of an interactive visualization dashboard

Uba Backonja <sup>1</sup>, Seungeun Park <sup>2</sup>, Amae Kurre <sup>3</sup>, Hayley Yudelman <sup>3</sup>, Sam Heindel <sup>3</sup>,  
Melinda Schultz <sup>4</sup>, Greg Whitman <sup>4</sup>, Anne M Turner <sup>5</sup>, Natasza T Marchak <sup>6</sup>, Betty Bekemeier <sup>7</sup>

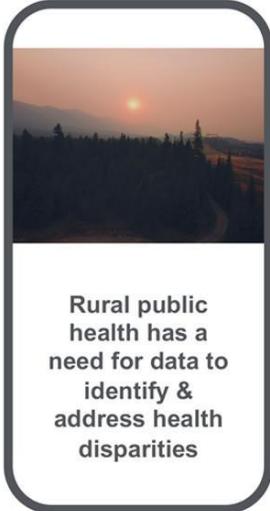
Affiliations + expand

PMID: 35318149 DOI: [10.1016/j.jbi.2022.104051](https://doi.org/10.1016/j.jbi.2022.104051)

# Supporting rural public health to address health disparities

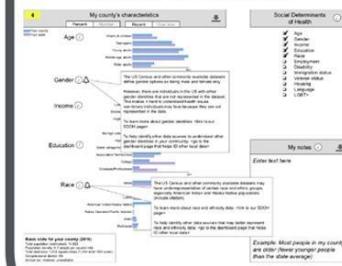


User-centered development of an interactive visualization dashboard for rural public health to address health disparities

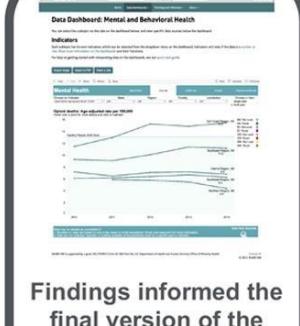
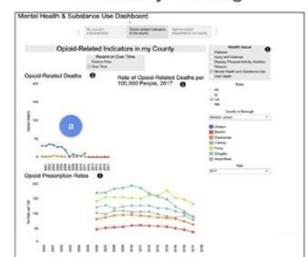


With end-users & stakeholders, we developed a website with dashboards for publicly available rural data via:

Mockup evaluations



Usability testing

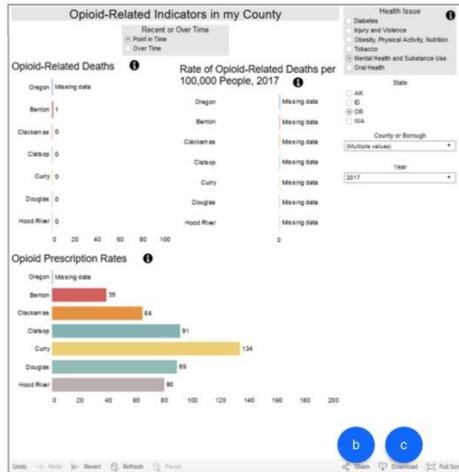
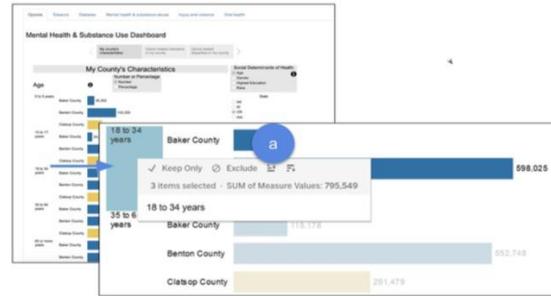


Funder: Office of the Assistant Secretary for Health (grant #1 CPIMP171144-01-00)

Backonja U, Park SE, Kure A, Yudelman H, Heindel S, Schultz M, Whitman G, Turner AM, Marchak NT, Bekemeier B. Supporting rural public health practice to address local-level social determinants of health across Northwest states: Development of an interactive visualization dashboard. *Journal of Biomedical Informatics*.

[sharenw.nwcphp.org](http://sharenw.nwcphp.org)

# Supporting rural public health to address health disparities



## Lesson learned example:

During usability testing of the alpha SHARE-NW website, participants had trouble finding (a) the **sort** function in the Tableau dashboard and (b) **the share and (c) download buttons**.

› NPJ Digit Med. 2022 Jun 29;5(1):81. doi: 10.1038/s41746-022-00623-8.

## International electronic health record-derived post-acute sequelae profiles of COVID-19 patients

Harrison G Zhang <sup># 1</sup>, Arianna Dagliati <sup># 2</sup>, Zahra Shakeri Hossein Abad <sup>1</sup>, Xin Xiong <sup>3</sup>,  
Clara-Lea Bonzel <sup>1</sup>, Zongqi Xia <sup>4</sup>, Bryce W Q Tan <sup>5</sup>, Paul Avillach <sup>1</sup>, Gabriel A Brat <sup>1</sup>,  
Chuan Hong <sup>1 6</sup>, Michele Morris <sup>7</sup>, Shyam Visweswaran <sup>7</sup>, Lav P Patel <sup>8</sup>, Alba Gutiérrez-Sacristán <sup>1</sup>,  
David A Hanauer <sup>9</sup>, John H Holmes <sup>10 11</sup>, Malarkodi Jebathilagam Samayamuthu <sup>7</sup>,  
Florence T Bourgeois <sup>12</sup>, Sehi L'Yi <sup>1</sup>, Sarah E Maidlow <sup>13</sup>, Bertrand Moal <sup>14</sup>, Shawn N Murphy <sup>15</sup>,  
Zachary H Strasser <sup>16</sup>, Antoine Neuraz <sup>17</sup>, Kee Yuan Ngiam <sup>18</sup>, Ne Hooi Will Loh <sup>19</sup>,  
Gilbert S Omenn <sup>20</sup>, Andrea Prunotto <sup>21</sup>, Lauren A Dalvin <sup>22</sup>, Jeffrey G Klann <sup>16</sup>, Petra Schubert <sup>23</sup>,  
Fernando J Sanz Vidorreta <sup>24</sup>, Vincent Benoit <sup>25</sup>, Guillaume Verdy <sup>14</sup>, Ramakanth Kavuluru <sup>26</sup>,  
Hossein Estiri <sup>16</sup>, Yuan Luo <sup>27</sup>, Alberto Malovini <sup>28</sup>, Valentina Tibollo <sup>28</sup>, Riccardo Bellazzi <sup>29</sup>,  
Kelly Cho <sup>23 30</sup>, Yuk-Lam Ho <sup>23</sup>, Amelia L M Tan <sup>1</sup>, Byorn W L Tan <sup>5</sup>, Nils Gehlenborg <sup>1</sup>,  
Sara Lozano-Zahonero <sup>21</sup>, Vianney Jouhet <sup>31</sup>, Luca Chiovato <sup>32</sup>, Bruce J Aronow <sup>33</sup>,  
Emma M S Toh <sup>34</sup>, Wei Gen Scott Wong <sup>35</sup>, Sara Pizzimenti <sup>36</sup>, Kavishwar B Wagholarikar <sup>16</sup>,  
Mauro Bucalo <sup>37</sup>; Consortium for Clinical Characterization of COVID-19 by EHR (4CE); Tianxi Cai <sup># 1</sup>,  
Andrew M South <sup># 38</sup>, Isaac S Kohane <sup># 1</sup>, Griffin M Weber <sup># 39</sup>

Affiliations + expand

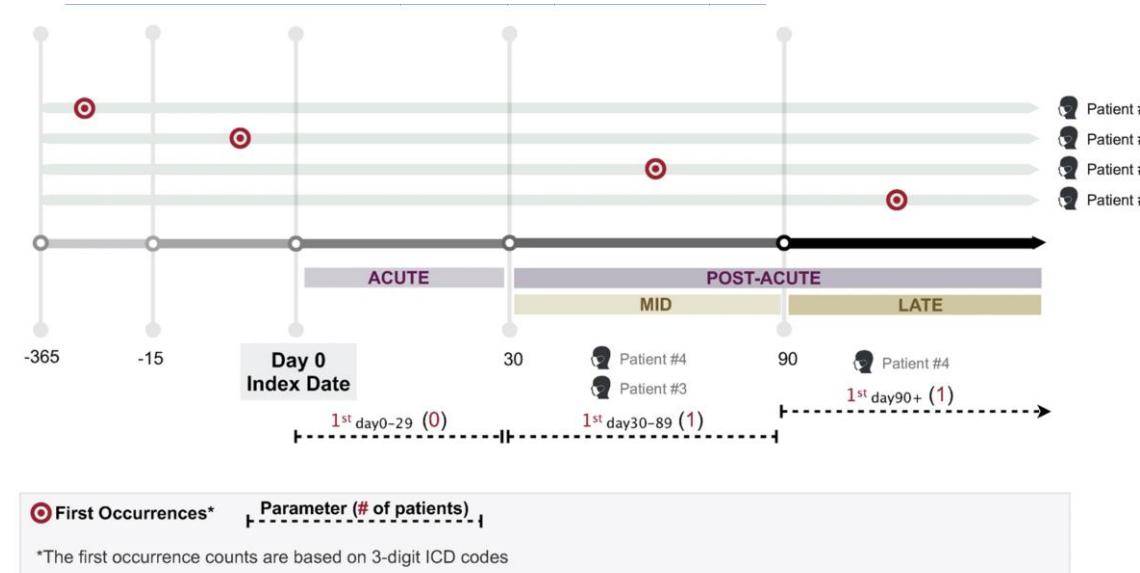
PMID: 35768548 PMCID: PMC9242995 DOI: 10.1038/s41746-022-00623-8

# International EHR-derived Long COVID profiles of COVID-19 patients

Identified incident annotations of phenotypes (use ICD codes) in a federated network of 277 international hospitals

Estimated relative risk of phenotypes among COVID-19+ patients (n=414,602) relative to non-infected patients (n=2,300,000) at:

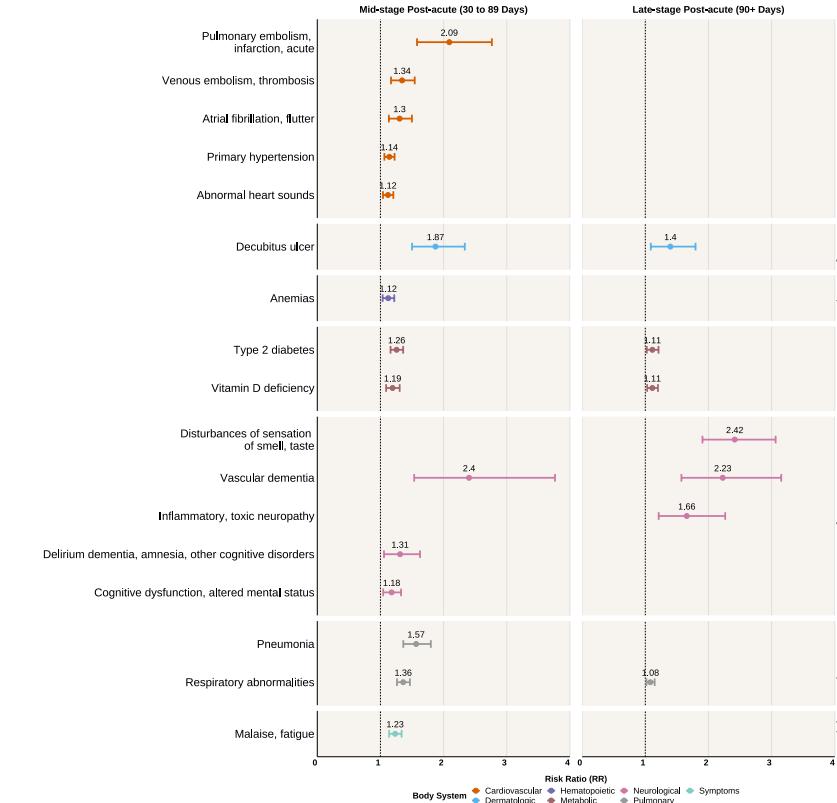
- 30 days after index date
- 90 days after index date



4CE

# Diseases with increased risk in COVID-19 outpatients

Using a federated international network, the authors systematically identified robust conditions associated with PASC compared to control groups, underscoring the multifaceted cardiovascular and neurological phenotype profiles of Long COVID



> Nat Med. 2023 Jan;29(1):226-235. doi: 10.1038/s41591-022-02116-3. Epub 2022 Dec 1.

## Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes

Hao Zhang <sup>1</sup>, Chengxi Zang <sup>1</sup>, Zhenxing Xu <sup>1</sup>, Yongkang Zhang <sup>1</sup>, Jie Xu <sup>2</sup>, Jiang Bian <sup>2</sup>,  
Dmitry Morozyuk <sup>1</sup>, Dhruv Khullar <sup>1</sup>, Yiye Zhang <sup>1</sup>, Anna S Nordvig <sup>3</sup>, Edward J Schenck <sup>4</sup>,  
Elizabeth A Shenkman <sup>2</sup>, Russell L Rothman <sup>5</sup>, Jason P Block <sup>6</sup>, Kristin Lyman <sup>7</sup>, Mark G Weiner <sup>1</sup>,  
Thomas W Carton <sup>7</sup>, Fei Wang <sup>8</sup>, Rainu Kaushal <sup>1</sup>

Affiliations + expand

PMID: 36456834 PMCID: PMC9873564 DOI: 10.1038/s41591-022-02116-3

# Identifying long-COVID subphenotypes

## DATABASE



Electronic Health Records (EHR) data for patients with lab-confirmed SARS-CoV-2 Infection from two clinical research networks (CRN)

- INSIGHT: New York City area
- OneFlorida+: Florida, Georgia and Alabama

## METHOD

Cohort: SARS-CoV-2 infected patients with newly incident conditions within 30-180 days after infection

Variables: 137 investigative conditions

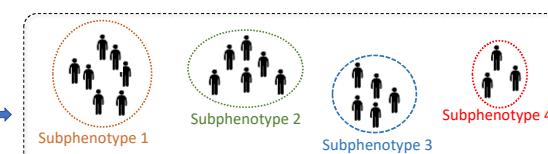
Step 1. Binary vector representations of patients with incident PASC diagnosis

Anemia	1	0	0	0	0
Circulatory problem	1	0	0	0	0
Disorders of stomach	0	1	0	0	0
Malaise and fatigue	0	1	1	0	1
Nausea and Vomiting	1	1	0	0	0
Headache	0	0	1	0	0
Respirator problem	0	0	1	0	0
...	...	...	...	...	...
Musculoskeletal pain	0	0	0	1	0
Osteoarthritis	0	0	0	1	0

Step 2. Inference of PASC topics. Each PASC topic encodes a particular set of frequently co-occurred incident PASC diagnosis

Topic 1	0.05	0.74	0.08	0.02
Topic 2	0.65	0.01	0.13	0.06
...	...	...	...	...
Topic K-1	0.12	0.03	0.53	0.03
Topic K	0.07	0.10	0.02	0.69

Step 3. Derivation of the patient representation in the PASC topic space



Step 4. Derivation of the PASC subphenotypes as patient groups with the PASC topic-based representation through cluster analysis.

## INSIGHT Database

INSIGHT Clinical Research Network till 2021  
N=5,346,357

Any SARS-CoV-2 PCR/Antigen tests, March 2020 to Nov. 2021  
N=1,402,348

With confirmed laboratory test results  
N=1,219,871

Adults with age $\geq$ 20  
N=1,070,073

Any diagnoses in 3 years to 7 days before COVID onset  
N=638,479

Any diagnoses in 30 days to 180 days after COVID onset  
N=384,373

SARS-CoV2 positive patients  
N=35,275

Have at least one PASC in 30 days to 180 days after COVID onset  
N=27,250

Eligible patients  
Covid19+ patients with new PASC  
N=20,881

Exclusion: Do not take any SARS-CoV-2 PCR/Antigen tests from March 2020 to Nov. 2021  
N=19,207,300

Exclusion: Without confirmed laboratory test results

Exclusion: age<20

Exclusion: Without any diagnoses in 3 years to 7 days before COVID onset

Exclusion: Without any diagnoses in 30 days to 180 days after COVID onset

Exclusion: Covid negative patients

Exclusion : Do not have any PASC in 30 days to 180 days after COVID onset

Exclusion: Patients with any of their PASC appearing in baseline period

## OneFlorida+ Database

OneFlorida+ Clinical Research Network till 2021  
N=19,207,300

Any SARS-CoV-2 PCR/Antigen tests, March 2020 to Nov. 2021  
N=718,199

With confirmed laboratory test results  
N=716,563

Adults with age $\geq$ 20  
N=612,104

Any diagnoses in 3 years to 7 days before COVID onset  
N=392,325

Any diagnoses in 30 days to 180 days after COVID onset  
N=212,403

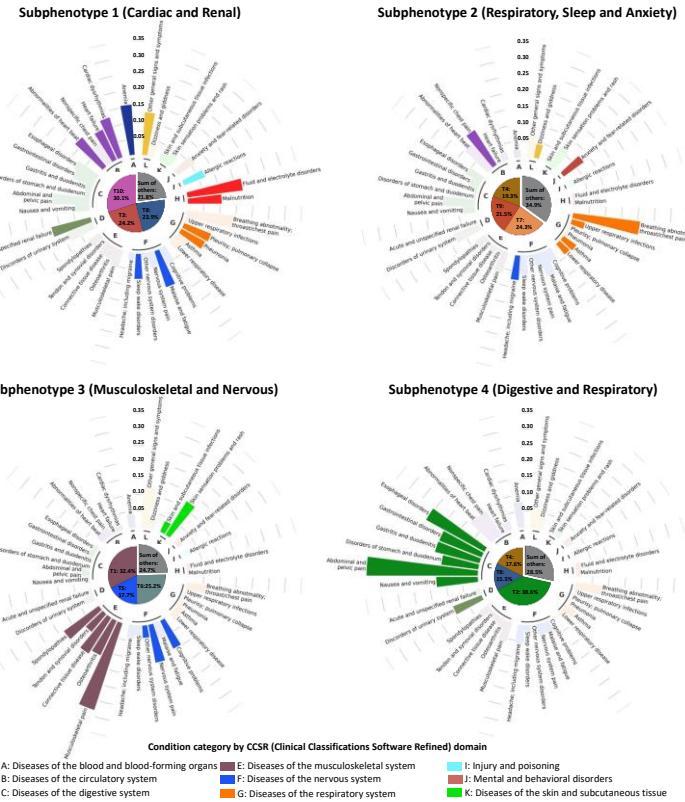
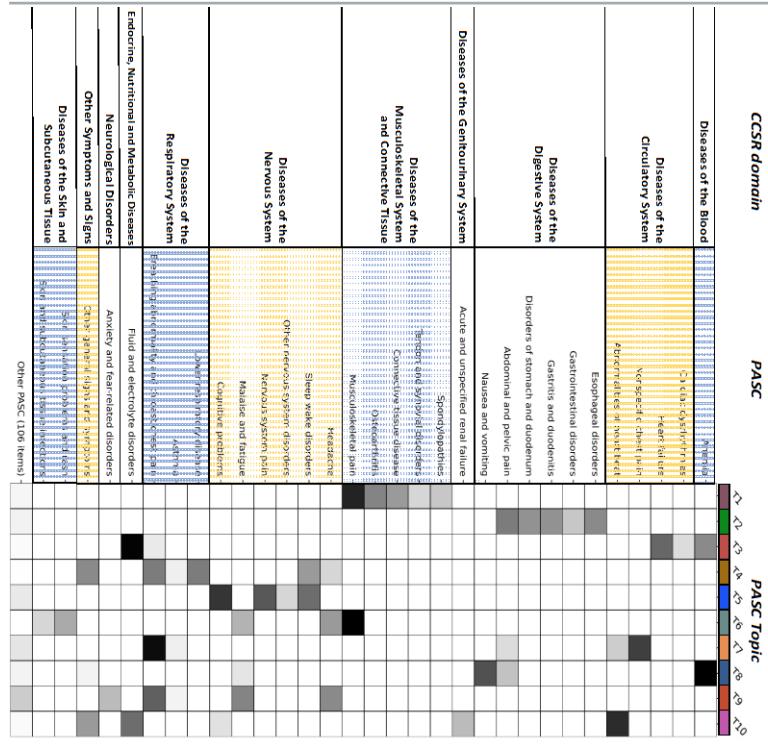
SARS-CoV2 positive patients  
N=22,341

Have at least one PASC in 30 days to 180 days after COVID onset  
N=18,434

Eligible patients  
Covid19+ patients with new PASC  
N=13,724

Zhang H et al, Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. Nat Med. 2023 Jan;29(1):226-235.

# Identifying long-COVID subphenotypes



> J Am Med Inform Assoc. 2022 Nov 14;29(12):2050-2056. doi: 10.1093/jamia/ocac178.

## Trading-off privacy and utility: the Washington State experience assessing the performance of a public health digital exposure notification system for coronavirus disease 2019

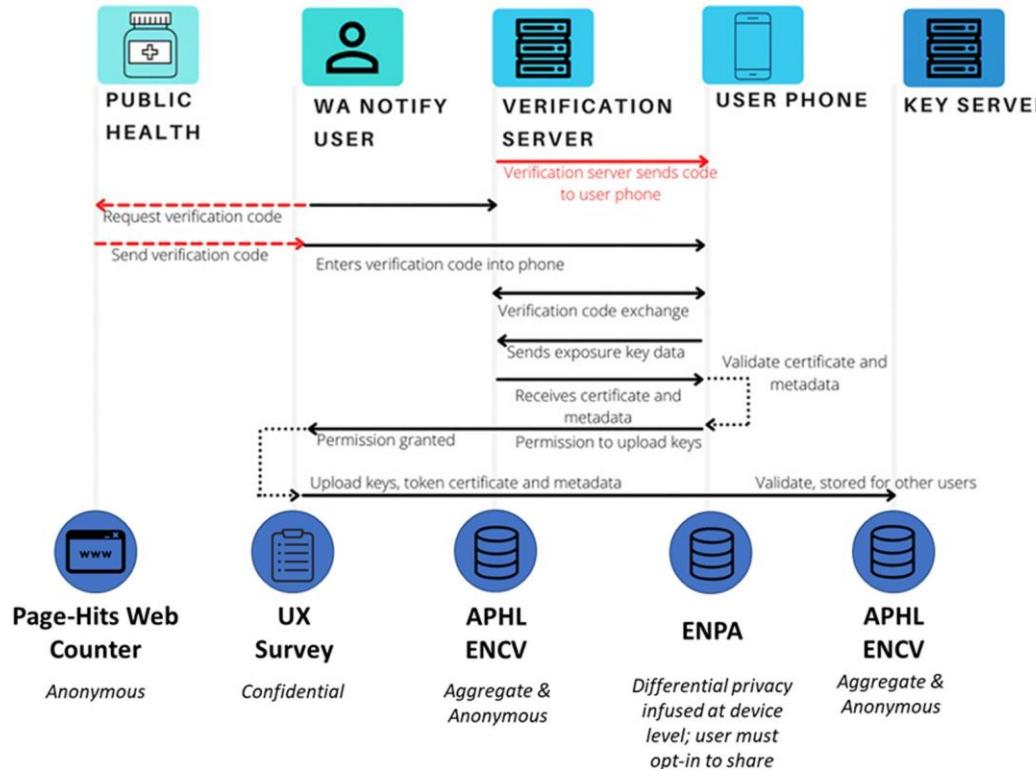
Courtney D Segal <sup>1</sup>, William B Lober <sup>2 3</sup>, Debra Revere <sup>1</sup>, Daniel Lorigan <sup>3</sup>, Bryant T Karras <sup>4</sup>,  
Janet G Baseman <sup>5</sup>

Affiliations + expand

PMID: 36206130 PMCID: PMC9667161 DOI: 10.1093/jamia/ocac178

# Digital Exposure Notifications

## Trading-off privacy and utility: the Washington State experience



# Digital Exposure Notifications

## Trading-off privacy and utility: the Washington State experience



Implementation Evaluation Construct	Performance Indicator	Metric	WA Notify Estimates as of March 2022
Reach	Number of installations within the population	Count of active and cumulative users measured by user installations	<b>&gt;3M activations</b>
Engagement	Utilization among exposed users	Count of ENs received, opened or dismissed	<b>ENs opened: &gt;742k EN webpage visits</b>
Effectiveness	Transmissions prevented	Model of cases averted	<b>5500 infections prevented during the first 4 months</b>
Effectiveness	Transmission characteristics	Adherence to quarantine reported in UX surveys	<b>53% of WA Notify user survey respondents reported staying home after receiving an EN</b>

## Mapping the global distribution of spotted fever group rickettsiae

### Methods



**BUGS!!!**



1. Searched literature about SFGR from PubMed, Web of Science, medRxiv, and bioRxiv published between 1906 and 2021.
2. Mapped global distributions of all confirmed species of SFGR detected in animals, vectors, and human beings.
3. Assessed ecological drivers for the distributions of 17 major SFGR species using machine learning algorithms, and mapped model-predicted risks.
4. Clustering SFGR with similar ecological niches and their spatial distribution.

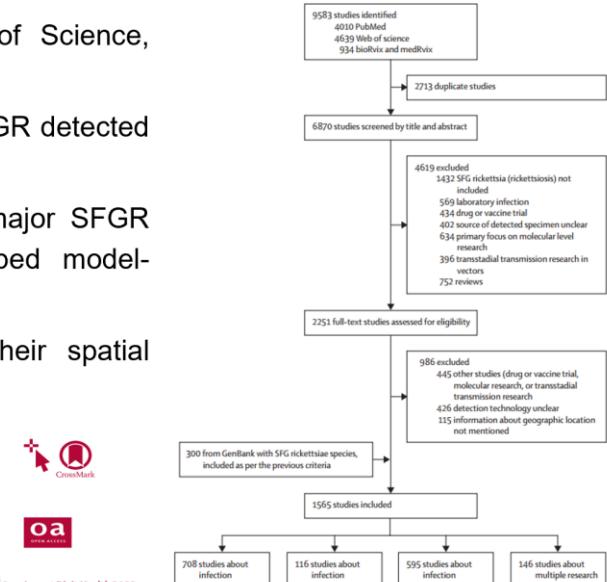
### Mapping the global distribution of spotted fever group rickettsiae: a systematic review with modelling analysis

Yuan-Yuan Zhang, Yan-Qun Sun, Jin-Jin Chen, Ai-Ying Teng, Tao Wang, Hao Li, Simon I Hay\*, Li-Qun Fang\*, Yang Yang\*, Wei Liu\*

#### Summary

**Background** Emerging and re-emerging spotted fever group (SFG) rickettsioses are increasingly recognised worldwide as threats to public health, yet their global distribution and associated risk burden remain poorly understood.

**Methods** In this systematic review and modelling analysis, we mapped global distributions of all confirmed species of SFG rickettsiae (SFGR) detected in animals, vectors, and human beings, using data collected from the literature. We



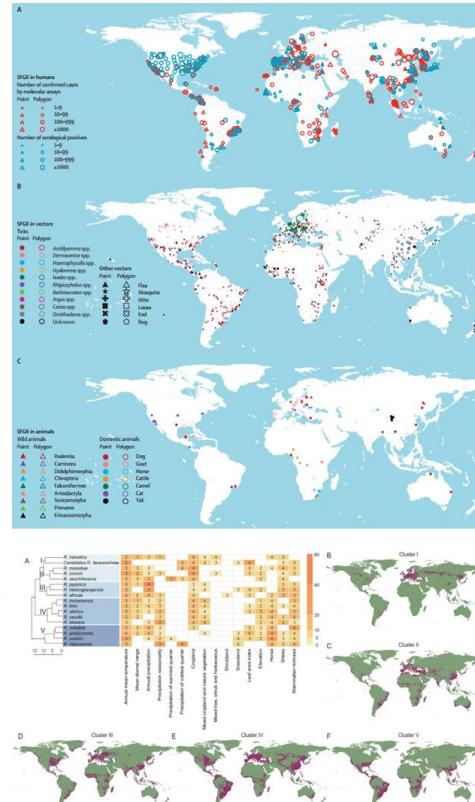
Lancet Digit Health 2023;  
5: e5–15  
Published Online  
November 21, 2022  
[https://doi.org/10.1016/S2589-7500\(22\)00212-6](https://doi.org/10.1016/S2589-7500(22)00212-6)

# Mapping the global distribution of spotted fever group rickettsiae

## Findings

1. 48 confirmed SFGR species, with **66133** human infections worldwide.
2. 198 vector species were detected to carry 47 of these *Rickettsia* spp. (146 ticks, 24 fleas, 15 mosquitoes, etc.).
3. 17 major SFGR species were found to **five** spatial clusters aggregated by ecological similarity.
4. *Rickettsia felis* is the leading SFGR species to which 4.4 billion people are at the suitable ecological areas of its occurrence, followed by *R. conorii* (3.7 billion) and *R. africae* (3.6 billion).
5. The most influential predictors of SFGR occurrence were climatic drivers, and some animal-related factors, including sheep density, horse density, and mammalian richness, contributed substantially as well

	Average testing AUC (2.5-97.5% percentiles)	Population size (million)					Area (10 000 km <sup>2</sup> )				
		Eurasia	Africa	Americas	Oceania	Worldwide	Eurasia	Africa	Americas	Oceania	Worldwide
<i>Rickettsia felis</i>	0.936 (0.910-0.955)	2936.3 (2513.9-3504.2)	720.7 (572.7-924.5)	764.8 (665.8-897.7)	22.7 (17.2-27.6)	4444.4 (3812.8-5338.3)	602.8 (463.2-778.1)	283.6 (207.7-420.1)	602.8 (479.4-761.4)	34.1 (24.9-48.4)	1523.3 (1203.3-2004.5)
<i>Rickettsia conorii</i>	0.949 (0.911-0.973)	2695.5 (2046.6-3283.3)	582.7 (442.9-711.8)	394.9 (255.9-513.4)	21.4 (15.1-25.0)	3694.5 (2789.9-4498.3)	570.2 (435.8-718.3)	232.7 (156.6-323.5)	251.0 (150.0-376.4)	67.3 (47.7-101.6)	1121.3 (775.8-1499.1)
<i>Rickettsia africae</i>	0.943 (0.882-0.973)	2355.3 (1727.6-3015.7)	998.0 (6917.1-1704.0)	250.8 (149.1-369.5)	14.4 (9.8-20.0)	3618.6 (2574.7-4426.9)	338.1 (220.6-502.1)	449.7 (278.6-628.1)	147.5 (75.2-277.8)	51.1 (31.6-90.1)	986.5 (625.8-1450.8)



# Ethical, Legal and Social Issues

# Ethical, Legal and Social Issues



› Nat Commun. 2022 Dec 9;13(1):7609. doi: 10.1038/s41467-022-35295-1.

## A Multifaceted benchmarking of synthetic electronic health record generation models

Chao Yan <sup># 1</sup>, Yao Yan <sup># 2</sup>, Zhiyu Wan <sup># 1</sup>, Ziqi Zhang <sup>3</sup>, Larsson Omberg <sup>2</sup>, Justin Guinney <sup>4 5</sup>, Sean D Mooney <sup>6</sup>, Bradley A Malin <sup>7 8 9</sup>

Affiliations + expand

PMID: 36494374 PMCID: PMC9734113 DOI: 10.1038/s41467-022-35295-1

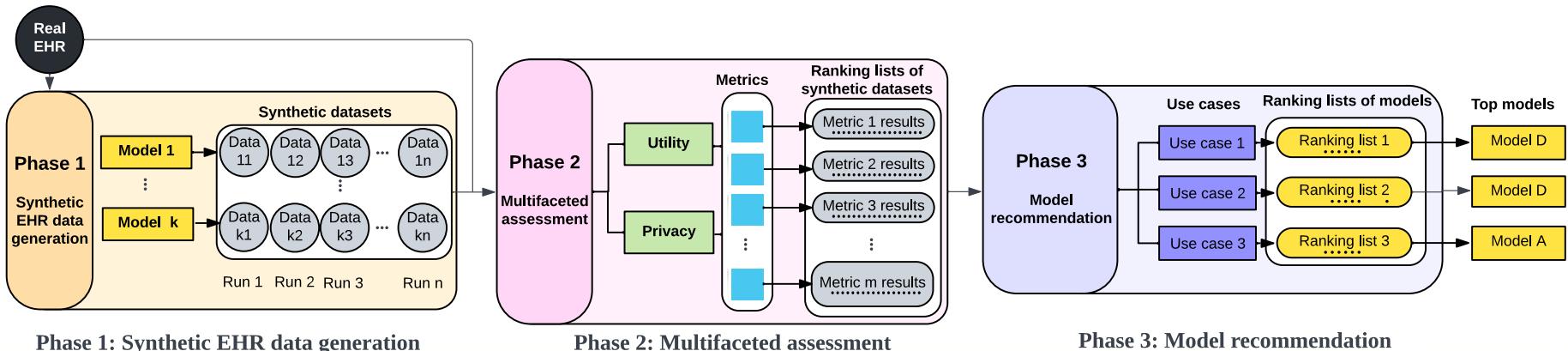
# Multifaceted benchmarking of synthetic EHR generation

## Synthetic EHR data generation:

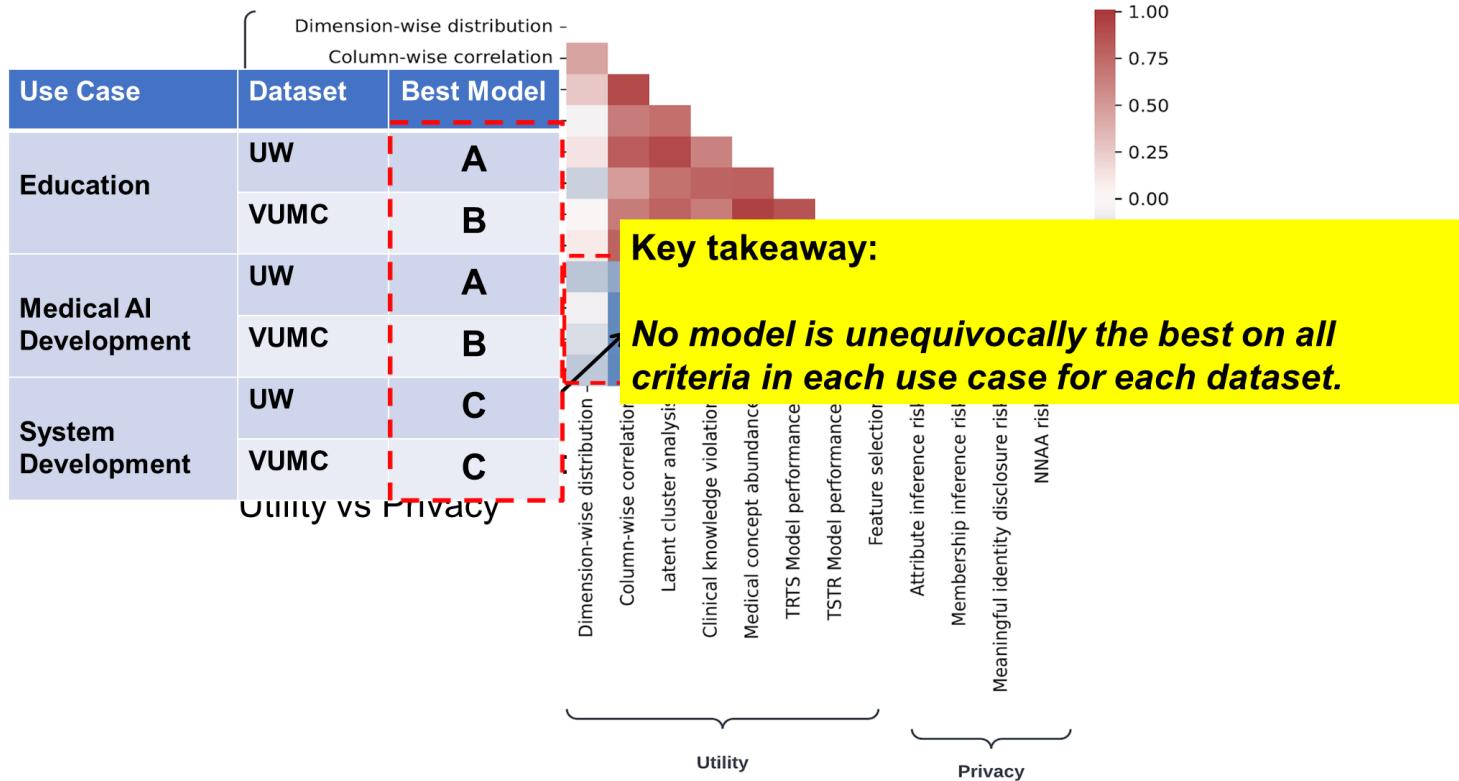
- A solution to broaden data sharing at scale
- Retain statistical nature of real data while protecting privacy

## Status quo:

- A lack of consensus on synthetic EHR data evaluation
- A disconnection between evaluation to use cases



# Multifaceted benchmarking of synthetic EHR generation



› J Biomed Inform. 2022 Aug;132:104113. doi: 10.1016/j.jbi.2022.104113. Epub 2022 Jun 9.

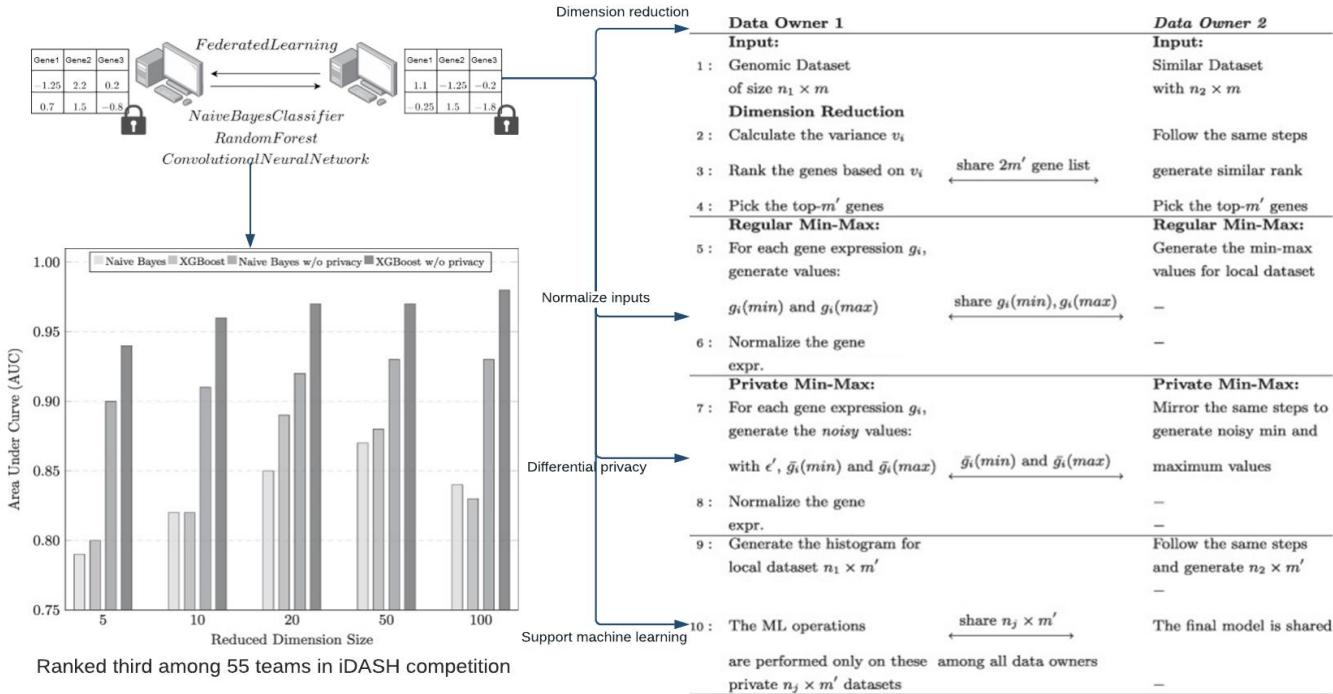
## Generalized genomic data sharing for differentially private federated learning

Md Momin Al Aziz <sup>1</sup>, Md Monowar Anjum <sup>2</sup>, Noman Mohammed <sup>2</sup>, Xiaoqian Jiang <sup>3</sup>

Affiliations + expand

PMID: 35690350 DOI: [10.1016/j.jbi.2022.104113](https://doi.org/10.1016/j.jbi.2022.104113)

# Private Federated Learning for Genomic Data



# Ethical, Legal and Social Issues



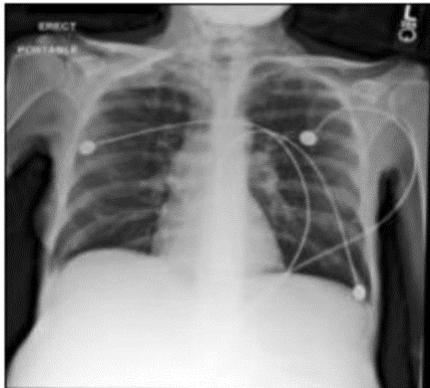
› Lancet Digit Health. 2022 Jun;4(6):e406-e414. doi: 10.1016/S2589-7500(22)00063-2.  
Epub 2022 May 11.

## AI recognition of patient race in medical imaging: a modelling study

Judy Wawira Gichoya <sup>1</sup>, Imon Banerjee <sup>2</sup>, Ananth Reddy Bhimireddy <sup>3</sup>, John L Burns <sup>4</sup>,  
Leo Anthony Celi <sup>5</sup>, Li-Ching Chen <sup>6</sup>, Ramon Correa <sup>2</sup>, Natalie Dullerud <sup>7</sup>, Marzyeh Ghassemi <sup>8</sup>,  
Shih-Cheng Huang <sup>9</sup>, Po-Chih Kuo <sup>6</sup>, Matthew P Lungren <sup>9</sup>, Lyle J Palmer <sup>10</sup>, Brandon J Price <sup>11</sup>,  
Saptarshi Purkayastha <sup>4</sup>, Ayis T Pyrros <sup>12</sup>, Lauren Oakden-Rayner <sup>13</sup>, Chima Okechukwu <sup>14</sup>,  
Laleh Seyyed-Kalantari <sup>15</sup>, Hari Trivedi <sup>3</sup>, Ryan Wang <sup>6</sup>, Zachary Zaiman <sup>16</sup>, Haoran Zhang <sup>7</sup>

Affiliations + expand

PMID: 35568690 PMCID: PMC9650160 DOI: 10.1016/S2589-7500(22)00063-2



# Pop Quiz!

# Is this patient Black?

Yes



No



- ① Start presenting to display the poll results on this slide.

# Pop Quiz!



Is this patient Black?

## Race detection in radiology imaging

### Chest x-ray (internal validation)\*

MXR (Resnet34, Densenet121)	0.97, 0.94
CXP (Resnet 34)	0.98
EMX (Resnet34, Densenet121, EfficientNet-B0)	0.98, 0.97, 0.99

### Chest x-ray (external validation)\*

MXR to CXP, MXR to EMX	0.97, 0.97
CXP to EMX, CXP to MXR	0.97, 0.96
EMX to MXR, EMX to CXP	0.98, 0.98

### Chest x-ray (comparison of models)†

MXR, CXP, EMX	Multiple results (appendix p 26)
---------------	----------------------------------

### CT chest (internal validation)\*

NLST (slice, study)	0.92, 0.96
---------------------	------------

### CT chest (external validation)\*

NLST to EM-CT (slice, study)	0.80, 0.87
NLST to RSPECT (slice, study)	0.83, 0.90

### Limb x-ray (internal validation)\*

DHA	0.91
-----	------

### Mammography\*

EM-Mammo (image, study)	0.78, 0.81
-------------------------	------------

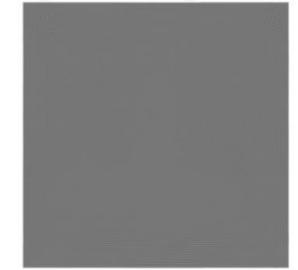
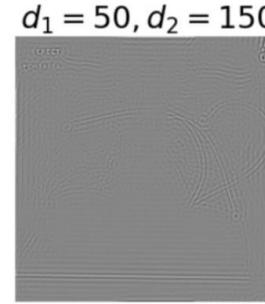
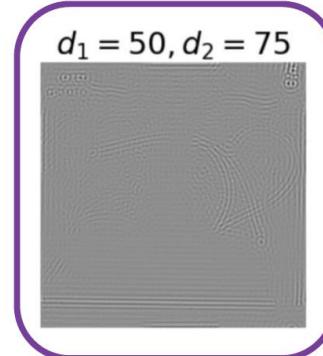
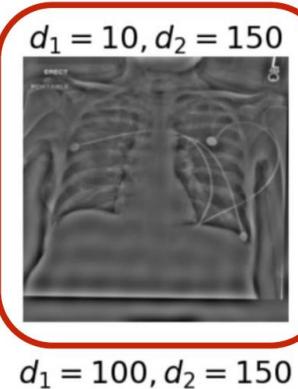
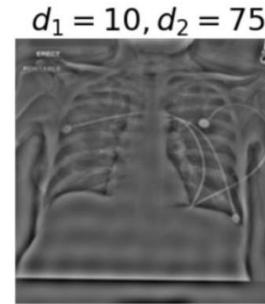
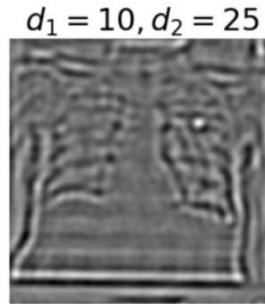
### Cervical spine x-ray\*

EM-CS	0.92
-------	------

# Self-reported Race is Obvious to AI

Race prediction AUROC

$d_1   d_2$	25	50	75	100	125	150
10	0.86	0.90	0.91	0.91	0.91	0.91
25		0.86	0.89	0.90	0.90	0.91
50			0.87	0.89	0.89	0.89
75				0.85	0.86	0.87
100					0.84	0.84
125						0.75



Gichoya, Judy W., et al. "AI recognition of patient race in medical imaging: a modelling study." Lancet Digital Health. 2022.

# Improving fairness in the prediction of HF patient outcomes



210,368 HF patients extracted from the Get With The Guidelines-Heart Failure (GWTG-HF) registry

Hospitalization LOS and in-hospital mortality

Stratified underdiagnosis and overdiagnosis rates

Circulation: Heart Failure

Volume 15, Issue 11, November 2022; Page e009473

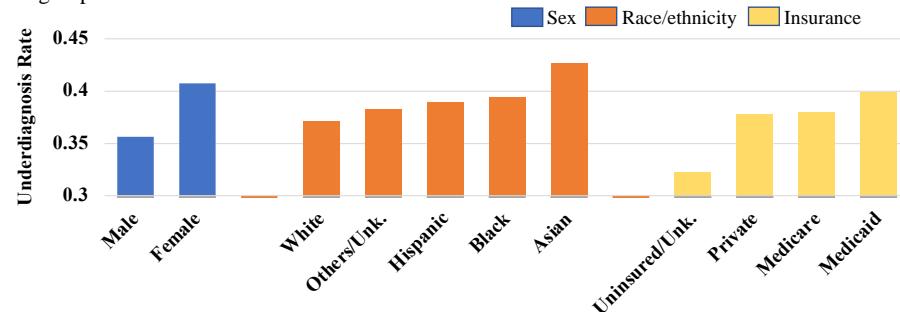


ORIGINAL ARTICLE

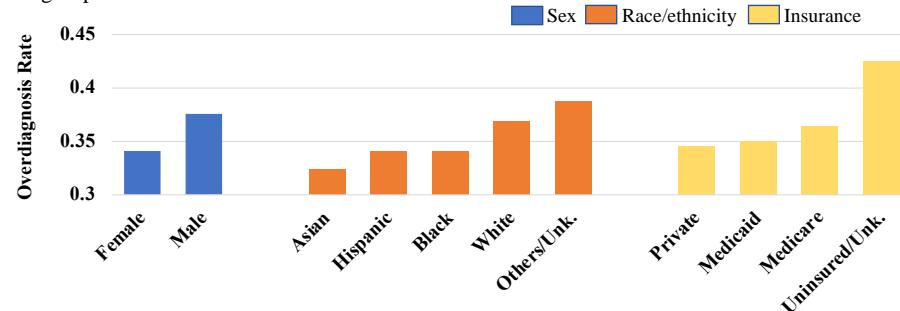
## Improving Fairness in the Prediction of Heart Failure Length of Stay and Mortality by Integrating Social Determinants of Health

Yikuan Li, MS , Hanyin Wang, BMED , and Yuan Luo, PhD

A. Underdiagnosis rates of the machine learning algorithms in sex, race/ethnicity and insurance subgroups



B. Overdiagnosis rates of the machine learning algorithms in sex, race/ethnicity and insurance subgroups



# Improving fairness in the prediction of HF patient outcomes

Variables *	Description	Domain
fpl_100	Percent population less than 100% federal poverty level	Income
sing_parent_fam	Percent single-parent households with dependents less than 18 years	Household
dropout	Percent drop-out (persons with no high school diploma estimate)	Education
no_car	Percent population with no car	Transportation
rent_occup	Percent renter occupied (tenure housing)	Housing
crowding	Percent crowded (tenure by occupants per room, greater than 1.01 to 1.50 occupants per room)	Housing
nonemp	Percent non-employed and not seeking work	Employment
unemp	Percent un-employed but actively seeking work	Employment
highneeds	Percent in high-needs age groups (children under the age of 5 and female between the ages of 15 and 44)	Demographics
hispanic	Percent Hispanic	Demographics
foreignb	Percent foreign born	Demographics
black	Percent non-Hispanic Black	Demographics
SDI	Social Deprivation Index	Comprehensive
ADI <sub>state</sub>	Area Deprivation Index - ranking at state level	Comprehensive
ADI <sub>national</sub>	Area Deprivation Index - ranking at national level	Comprehensive
all SDOH	Integration of all SDOH variables above	Collective

\* The abbreviated variable names were inherited from the original source of Social Deprivation Index database.

Integrated Variables*	Fairness	Performance			
	Demographic parity ratio	Equalized odds ratio	AUROC	Recall	p-Value
Baseline	0.828	0.826	0.680	0.654	1.000
fpl_100	<u>0.851</u>	<u>0.845</u>	0.682	0.656	0.952
sing_parent_fam	0.821	0.821	0.681	0.651	0.076
dropout	<u>0.865</u>	<u>0.864</u>	0.681	0.654	0.201
no_car	<u>0.835</u>	<u>0.821</u>	0.682	0.655	0.545
rent_occup	<u>0.844</u>	<u>0.851</u>	0.682	<b>0.657</b>	0.484
crowding	<u>0.873</u>	<u>0.872</u>	0.682	0.654	0.856
nonemp	<u>0.833</u>	<u>0.831</u>	0.681	0.655	0.349
unemp	<u>0.841</u>	<u>0.838</u>	0.681	0.656	0.951
highneeds	<u>0.852</u>	<u>0.857</u>	0.682	<b>0.657</b>	1.000
hispanic	<u>0.848</u>	<u>0.851</u>	<b>0.683</b>	<b>0.657</b>	0.178
foreignb	<u>0.845</u>	<u>0.845</u>	<b>0.683</b>	0.655	0.114
black	<u>0.866</u>	<u>0.885</u>	0.682	0.653	0.551
SDI	<u>0.855</u>	<u>0.865</u>	0.680	0.654	0.879
ADI <sub>national</sub>	<u>0.850</u>	<u>0.857</u>	0.681	0.653	0.220
ADI <sub>state</sub>	<u>0.830</u>	<u>0.829</u>	0.682	0.653	0.366
all SDOH	<b>0.881</b>	<u>0.863</u>	0.681	0.654	0.071

\* The abbreviated variable names were inherited from the original source of Social Deprivation Index database. For a detailed description of these variables, please refer Table 1.

> Sci Rep. 2022 Jul 22;12(1):12542. doi: 10.1038/s41598-022-16615-3.

## Prediction performance and fairness heterogeneity in cardiovascular risk models

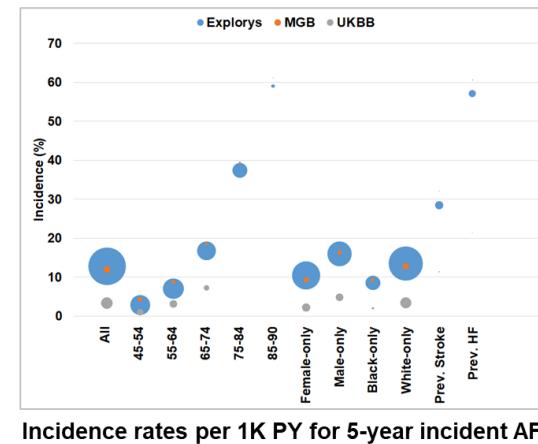
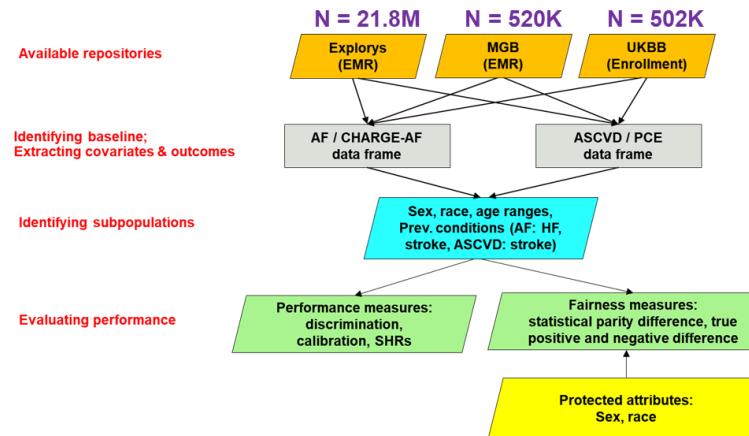
Uri Kartoun # 1, Shaan Khurshid # 2 3, Bum Chul Kwon 1, Aniruddh P Patel 2 4, Puneet Batra 5, Anthony Philippakis 2, Amit V Khera 2 4, Patrick T Ellinor 2 3, Steven A Lubitz 2 3, Kenney Ng 6

Affiliations + expand

PMID: 35869152 PMCID: PMC9307639 DOI: 10.1038/s41598-022-16615-3

# Heterogeneity in Cardiovascular Risk Prediction

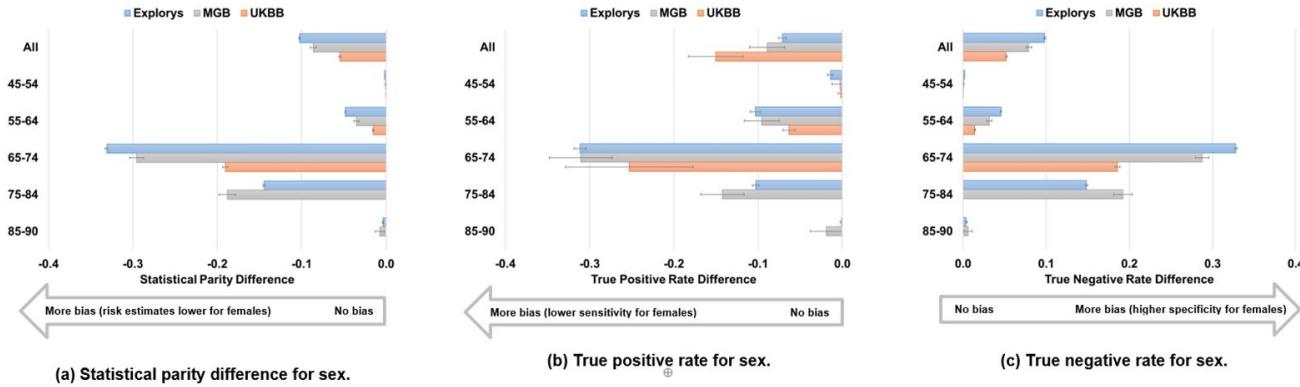
- The study is the first to report on fairness-related measures for the CHARGE-AF (to predict 5-year incident AF) and PCE (to predict 10-year incident ASCVD) scores to assess possible biases considering sex and race differences.
- Analysis of three large independent datasets identified important patterns of performance heterogeneity across clinically relevant subgroups.
- The results highlight the magnitude of poor performance affecting a large proportion of the population (discrimination, calibration, or both), especially patients at older ages and with prevalent conditions.



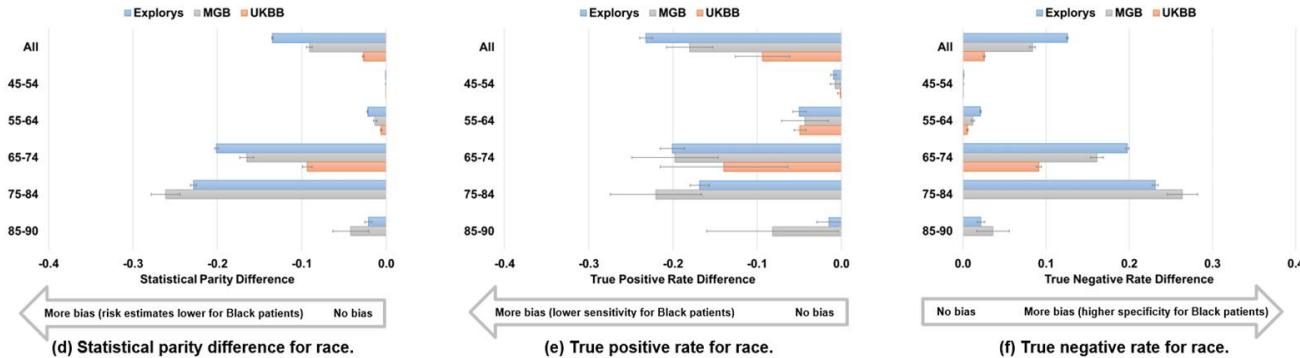
Kartoun, U., Khurshid, S., Kwon, B.C. et al. Prediction performance and fairness heterogeneity in cardiovascular risk models. Sci Rep 12, 12542 (2022).

# Heterogeneity in Fairness / CHARGE-AF

Sex



Race



Kartoun, U., Khurshid, S., Kwon, B.C. et al. Prediction performance and fairness heterogeneity in cardiovascular risk models. Sci Rep 12, 12542 (2022).

> Nat Commun. 2022 Nov 19;13(1):7094. doi: 10.1038/s41467-022-34592-z.

## Disparate impacts on online information access during the Covid-19 pandemic

Jina Suh <sup>1</sup> <sup>2</sup>, Eric Horvitz <sup>1</sup> <sup>2</sup>, Ryen W White <sup>1</sup> <sup>2</sup>, Tim Althoff <sup>3</sup>

Affiliations + expand

PMID: 36402817 PMCID: PMC9675823 DOI: 10.1038/s41467-022-34592-z

# Disparate Impacts on Online Information Access during the COVID-19 Pandemic

## DURING THE PANDEMIC

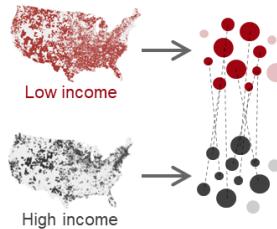
Digital engagement became a **necessity**.  
Health disparities were exacerbated.  
How about digital disparities?



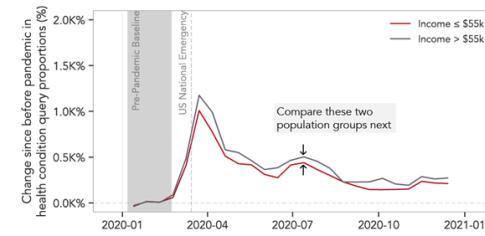
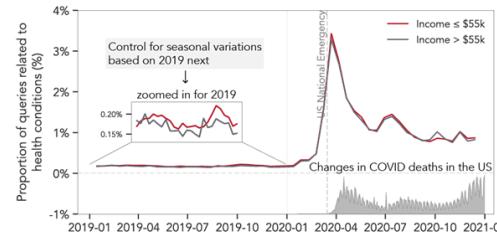
How did the shifts in online information access differ across socioeconomic and environmental factors?

## FOR EXAMPLE

How did the shifts in online health condition query behavior differ across high- and low-income ZIP code groups?



**55b**  
Bing search interactions  
across  
**25k**  
ZIP codes  
covering  
**97.2%**  
of US population



### 1. Match ZIP code groups

Create high- and low-income ZIP code groups that are comparable across socioeconomic and environmental factors

### 2. Measure changes

From 55b Bing search interactions, compute changes in the proportion of search queries relevant to health conditions, accounting for seasonal and temporal variations and pre-pandemic baselines.

### 3. Compare groups

Compare the changes in query proportions between two income ZIP code groups.

## 8 Socioeconomic and environmental factors used for comparing ZIP codes and matching:

Income, insurance access, education access, employment, internet access, population density, racial/ethnic distribution

## 7 Outcomes examined in the analysis:

Health condition queries, online learning site visits, unemployment queries, unemployment site visits, financial assistance queries, online food delivery queries, food assistance queries

## IMPLICATIONS

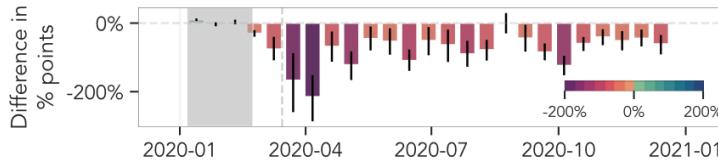
Web search data can be used for monitoring digital disparities **in real time and at population scales**.

Important for **downstream real-world implications** of differential information search behaviors.

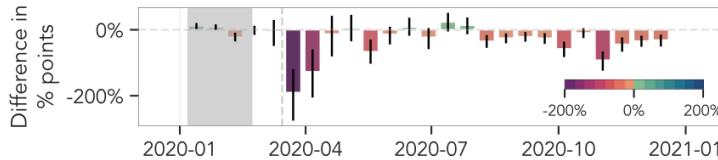
Isolation of factors through matching helps identify **potential barriers** to accessing digital resources or **disproportionately placed burdens** as well as highlight opportunities for **factor-specific interventions**.

# Disparate Impacts on Online Information Access during the COVID-19 Pandemic

Some communities were **not able to ramp up** their utilization of online resources as much.

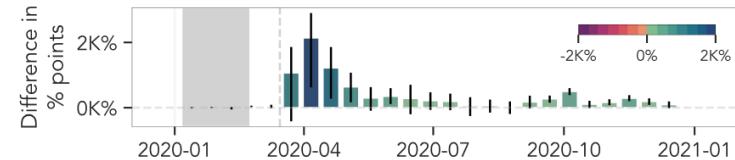


**Health condition queries** income  $\leq \$55k$  minus income  $> \$55k$   
ZIP codes associated with lower income show over a 200 percentage point smaller increase in health condition queries compared to high income group around April 2020.

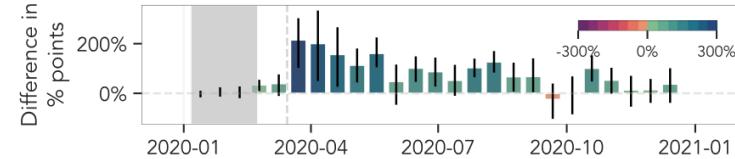


**Online learning site visits** income  $\leq \$55k$  minus income  $> \$55k$   
ZIP codes with lower income exhibited almost 200 percentage point smaller increase in online learning site visits in comparison around April 2020.

Some communities **supercharged** their utilization of online resources in comparison.



**Unemployment queries** % Black residents  $\geq 12\%$  minus % Black  $< 12\%$   
ZIP codes with higher proportions of Black residents show almost 2000 percentage point higher increase in unemployment search queries compared to their counterpart groups around April 2020.



**Food assistance queries** % attained BA  $\leq 21\%$  minus % BA  $> 21\%$   
ZIP codes with lower educational attainment exhibited around 200 percentage point higher increase in food assistance queries than their counterpart groups around April 2020.

# Clinical Decision Support

# Clinical Decision Support



Multicenter Study > BMJ. 2022 Feb 17;376:e068576. doi: 10.1136/bmj-2021-068576.

## Early identification of patients admitted to hospital for covid-19 at risk of clinical deterioration: model development and multisite external validation study



Fahad Kamran <sup>1 2</sup>, Shengpu Tang <sup>1 2</sup>, Erkin Otles <sup>3 4</sup>, Dustin S McEvoy <sup>5</sup>,  
Sameh N Saleh <sup>6 7</sup>, Jen Gong <sup>8</sup>, Benjamin Y Li <sup>1 4</sup>, Sayon Dutta <sup>5 9</sup>, Xinran Liu <sup>10</sup>,  
Richard J Medford <sup>6 7</sup>, Thomas S Valley <sup>11 12</sup>, Lauren R West <sup>13</sup>, Karandeep Singh <sup>11 14</sup>,  
Seth Blumberg <sup>10 15</sup>, John P Donnelly <sup>11 14</sup>, Erica S Shenoy <sup>13 16 17</sup>,  
John Z Ayanian <sup>11 12</sup>, Brahmajee K Nallamothu <sup>11 12</sup>, Michael W Sjoding <sup>11 12 18</sup>,  
Jenna Wiens <sup>19 11 18</sup>

Affiliations + expand

PMID: 35177406 PMCID: PMC8850910 DOI: 10.1136/bmj-2021-068576

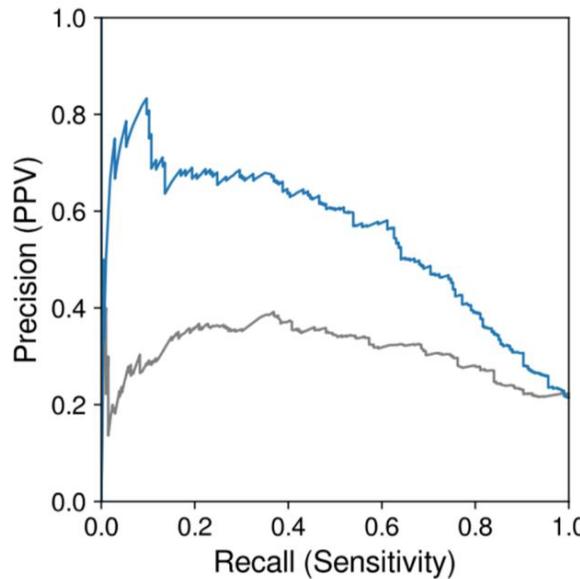
# Triaging ‘high risk’



Can we identify patients with covid-19 at risk of clinical deterioration?

Task: Identify individuals who will experience an adverse outcome within the first 5 days of admission

Validation cohort: 956 (21.6%)



## M-CURES

AUC .80 (95% CI .77, .84)  
Sensitivity: 48.8%  
Positive Pred. Value: 65%

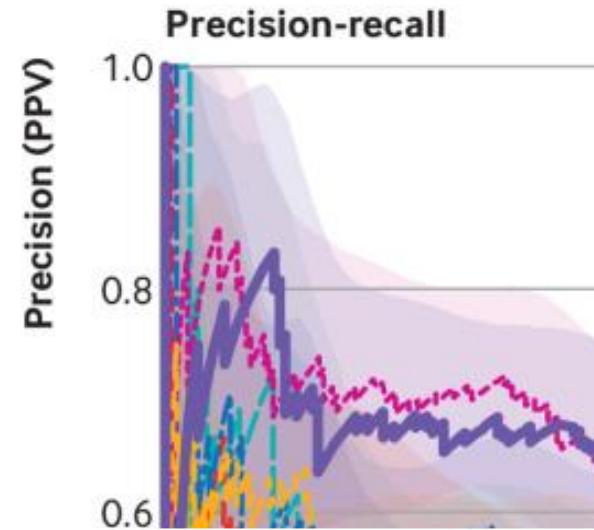
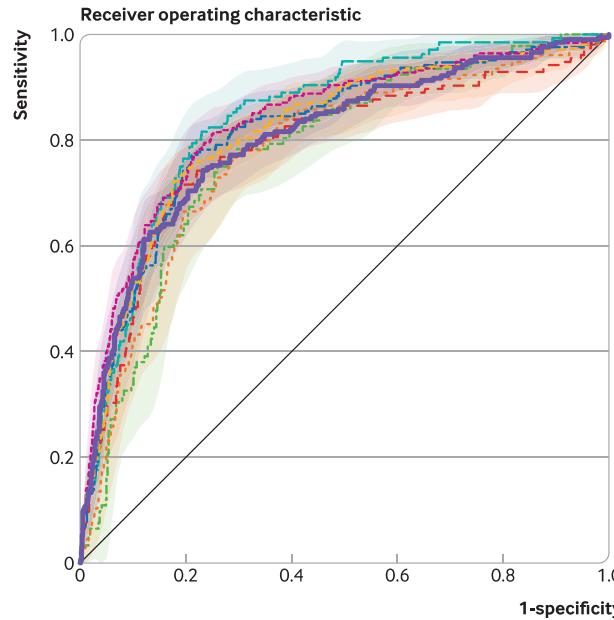
## EPIC-DI

AUC .66 (95% CI .62, .70)  
Sensitivity: N/A  
Positive Pred. Value: 65%

*nine clinical and personal characteristic variables selected from 2686 variables*

# External Validation

validated on over 9,000 COVID-19 positive individuals at Michigan Medicine and 12 other external hospitals.



# Building Risk Stratification Models that Generalize



Leverage data from related tasks when data are limited – trained on a related cohort of individuals presenting in the emergency department with shortness of breath

Clinician-in-the-loop validation – small number of COVID-19 cases on which to validate, scrutinize the model and ensure that it aligns with clinical intuition

Share code when you cannot share data – to assess generalizability worked with other hospital centers to share code and test model on different patient populations

# Clinical Decision Support



› Nat Med. 2022 Jul;28(7):1455-1460. doi: 10.1038/s41591-022-01894-0. Epub 2022 Jul 21.

## Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis



Roy Adams <sup>1 2</sup>, Katharine E Henry <sup>2 3</sup>, Anirudh Sridharan <sup>4</sup>, Hossein Soleimani <sup>5</sup>,  
Andong Zhan <sup>2 3</sup>, Nishi Rawat <sup>6</sup>, Lauren Johnson <sup>7</sup>, David N Hager <sup>8</sup>,  
Sara E Cosgrove <sup>8</sup>, Andrew Markowski <sup>9</sup>, Eili Y Klein <sup>10</sup>, Edward S Chen <sup>8</sup>,  
Mustapha O Saheed <sup>10</sup>, Maureen Henley <sup>7</sup>, Sheila Miranda <sup>11</sup>, Katrina Houston <sup>7</sup>,  
Robert C Linton <sup>4</sup>, Anushree R Ahluwalia <sup>7</sup>, Albert W Wu <sup>12 13 14 15 16</sup>, Suchi Saria <sup>17 18 19 20 21</sup>

Affiliations + expand

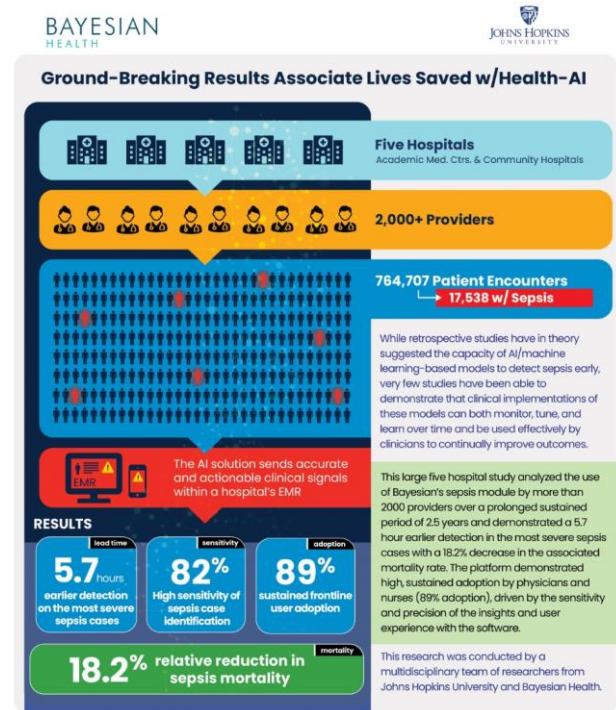
PMID: 35864252 DOI: 10.1038/s41591-022-01894-0

# Targeted Real-time Early Warning System (TREWS)

**Sepsis** is one of the leading causes of in-hospital mortality and **early identification and treatment** key to successful outcomes.

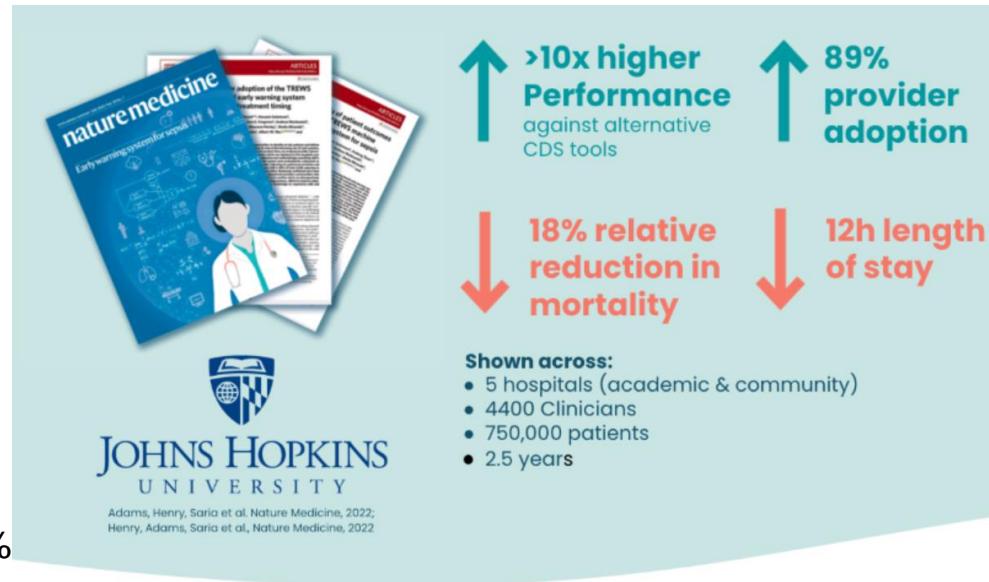
The Targeted Real-time Early Warning System (TREWS) for sepsis uses machine learning to detect early sepsis symptoms

- Delivers **bedside** alerts
- Deployed enterprise-wide across **five Johns Hopkins hospitals** in partnership with Bayesian Health (a Health AI platform company)
- Results reflect data from **over 500k screened patients**



# Patient outcomes

- Confirmation of TREWS within 3 hours also associated with improvements in sepsis outcomes:
  - In-hospital mortality: 3.3% absolute and 18.2% relative reduction
  - Median length of stay: 11.5 hour reduction
  - Organ failure: 0.26 point reduction in average SOFA score
- Associations larger among high-risk patients
  - In-hospital mortality: 4.5% absolute and 13% relative reduction
  - Median length of stay: 14.2 hour reduction
  - Organ failure: 0.38 point reduction in average SOFA score



# Provider adoption of TREWS

## Predictive performance and adoption:

- Providers responded to **89% of alerts** and 73% within 3 hours
- 38% of alerts were confirmed in real-time (precision)
- 82% of sepsis cases were flagged by TREWS (sensitivity)

**Impact on patient care:** Confirmation of TREWS within 3 hours associated with **1.81 hour reduction in median time to first antibiotics** among sepsis patients

- Critical process of care outcome for sepsis
- Adjusted for patient, provider, and environmental factors

*Henry, Adams, ... Saria, Nature Medicine (2022)*

In interviews, users of the system:

- Did not draw a hard distinction between ML and rule-based tools
- Perceived themselves as **partnering with the tool**
- Valued having the tool be supportive as opposed to taking autonomy away

*Henry, ... Mutlu, Saria, npj Digital Medicine (2022)*

› J Am Med Inform Assoc. 2022 Apr 13;29(5):805-812. doi: 10.1093/jamia/ocac023.

## Documentation of hospitalization risk factors in electronic health records (EHRs): a qualitative study with home healthcare clinicians



Mollie Hobensack <sup>1</sup>, Marietta Ojo <sup>2</sup>, Yolanda Barrón <sup>2</sup>, Kathryn H Bowles <sup>2 3</sup>, Kenrick Cato <sup>1 4</sup>, Sena Chae <sup>5</sup>, Erin Kennedy <sup>3</sup>, Margaret V McDonald <sup>2</sup>, Sarah Collins Rossetti <sup>1 6</sup>, Jiyoun Song <sup>1 2</sup>, Sridevi Sridharan <sup>2</sup>, Maxim Topaz <sup>1 2 7</sup>

Affiliations + expand

PMID: 35196369 PMCID: PMC9006696 DOI: 10.1093/jamia/ocac023

# EHR Documentation for CDS



## EHR Documentation

drives clinical decision support (CDS) to help prevent hospitalizations

## To build a CDS, we aimed to understand:

- 1 What risk factors HHC clinicians associated with higher hospitalization risk
- 2 How to extract this information from the EHR

## Qualitative Interview Results:



**15** interdisciplinary clinicians interviewed

**79**

Risk factors were mapped to standardized terminologies (e.g., Omaha System, SNOMED-CT)



HHC clinicians stated that a majority of risk factors can be found in **clinical notes**



Full manuscript:

# Clinical Decision Support



› Lancet Digit Health. 2022 Jun;4(6):e426-e435. doi: 10.1016/S2589-7500(22)00041-3.

## Development and multimodal validation of a substance misuse algorithm for referral to treatment using artificial intelligence (SMART-AI): a retrospective deep learning study



Majid Afshar <sup>1</sup>, Brihat Sharma <sup>2</sup>, Dmitriy Dligach <sup>3</sup>, Madeline Oguss <sup>4</sup>, Randall Brown <sup>5</sup>,  
Neeraj Chhabra <sup>6</sup>, Hale M Thompson <sup>2</sup>, Talar Markossian <sup>7</sup>, Cara Joyce <sup>8</sup>,  
Matthew M Churpek <sup>4</sup>, Niranjan S Karnik <sup>2</sup>

Affiliations + expand

PMID: 35623797 PMCID: PMC9159760 DOI: 10.1016/S2589-7500(22)00041-3

# Deep Learning for Substance Use

**Problem:** No formal screening program at most hospitals despite recommendation from US Preventive Services Taskforce

**Solution:** Leveraging data collected during usual care clinical documentation for automated screening

**Inclusion Cohort:** All Hospitalized Adults

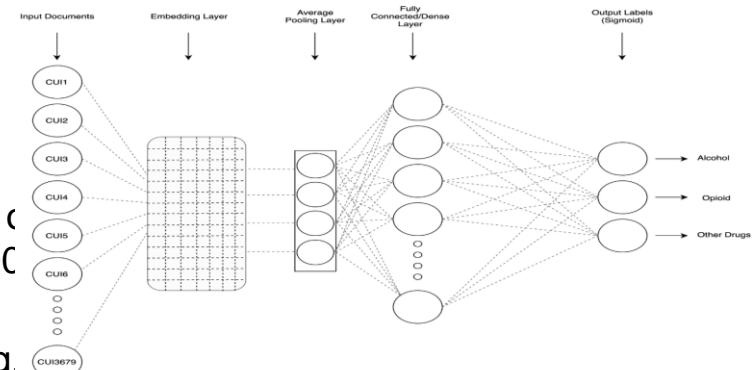
**Training Data:** 1<sup>st</sup> 24 hours of all notes and reports in the EHR

**Training Labels:** Manual screening with validated questionnaire for substance misuse (alcohol, opioid, and non-alcohol) in over 60,000 patients

**Models Evaluated:** Keyword rule-based, linear machine learning, non-linear deep learning

**Validation:** Internal temporal validation in prospective cohort; External validation at another health system

Best model was a Multilabel CNN using concept embeddings as features



Trained Model at <https://github.com/Rush-SubstanceUse-AILab/SMART-AI>



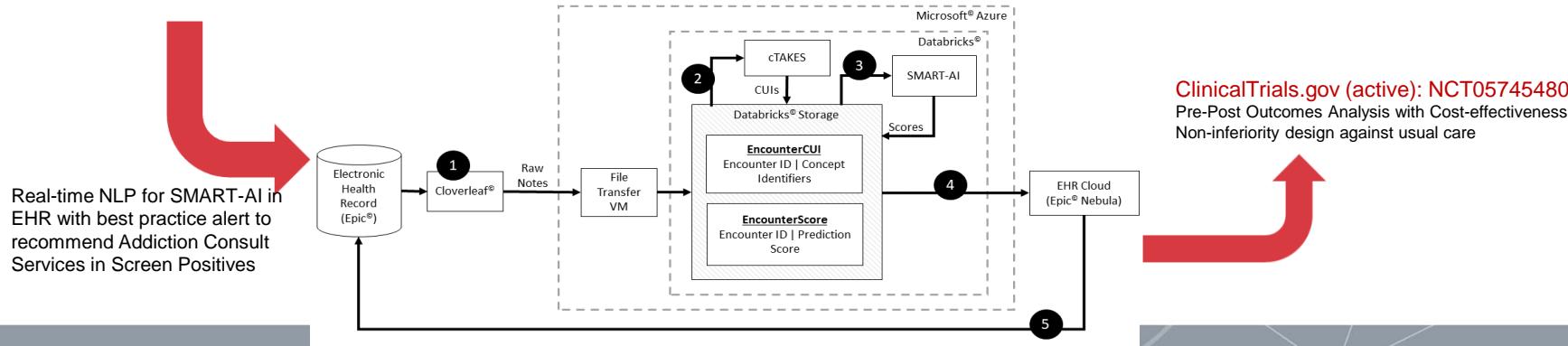
# Results and Implementation for Clinical Study

**Discrimination and Alarm Rate:** i) Overall AUROC: 0.97 (95% CI: 0.96 - 0.98); ii) Overall AUPRC: 0.69 (95% CI: 0.64 - 0.74); iii) Number Needed to Evaluate: 1.5 for alcohol misuse, 1.3 for opioid misuse, 2.6 for non-opioid drug misuse; and iv) 39, 26, and 16 alerts per 1000 patients for each group

**Calibration:** Well calibrated with slope and intercept near 0 and 1

**Bias and Fairness:** i) False omission rate and the false-positive rate remained low at 3% or less across subgroups of age, sex, and race and ethnicity; ii) False-negative rates and false-positive rates were similar between non-Hispanic Black and non-Hispanic White groups.

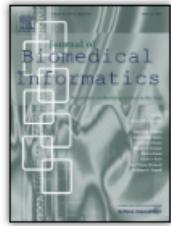
**Explainability/Face Validity:** Top features were ethanol for alcohol misuse, heroin for opioid misuse, and cocaine for non-opioid drug misuse



# Consumer and Pervasive Health Informatics

> J Biomed Inform. 2022 Apr;128:104040. doi: 10.1016/j.jbi.2022.104040. Epub 2022 Mar 6.

## Question-aware transformer models for consumer health question summarization



Shweta Yadav <sup>1</sup>, Deepak Gupta <sup>2</sup>, Asma Ben Abacha <sup>3</sup>, Dina Demner-Fushman <sup>4</sup>

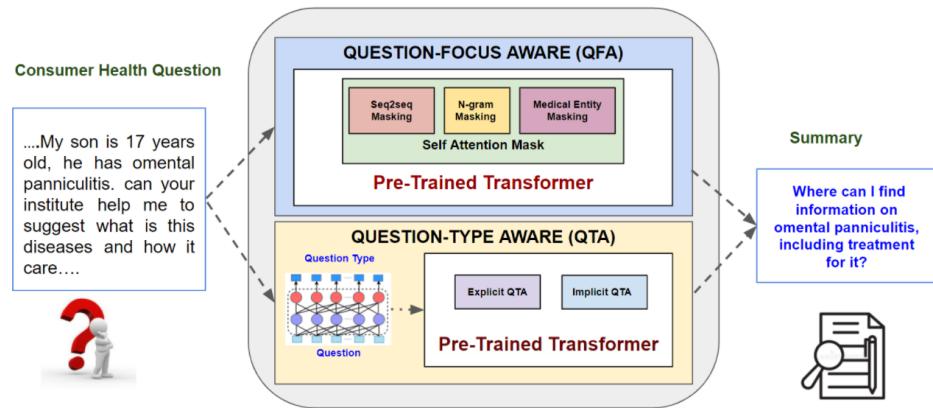
Affiliations + expand

PMID: 35259544 DOI: 10.1016/j.jbi.2022.104040



# Question-aware Transformer Models for Consumer Health Question Summarization

## Methods



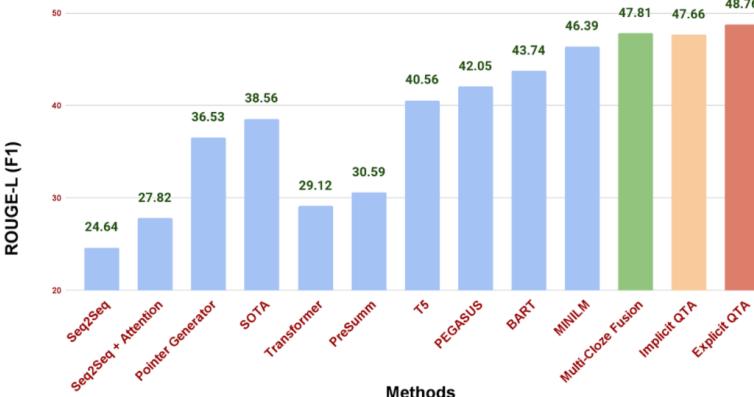
## Conclusion

Proposed question-focus and question-type aware summarization models that are able to generate relevant and succinct summaries of the original questions.

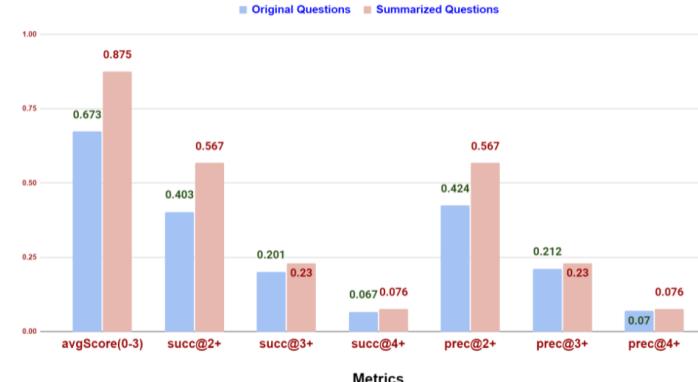
Proposed model achieves state-of-the-art performance on the MeQSum dataset outperforming various encoder-decoder and existing pretrained transformer-based methods.

## Results

Performance comparison of the proposed methods and various baseline models



Performance comparison of answer retrieval using original and automatically summarized questions



# Contributions



MeQSum benchmark of 1,000 consumer health questions and their associated summaries.

Introduce various Cloze tasks to pretrained transformer models for better coverage of the question focus

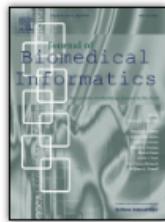
Introduce explicit and implicit ways of infusing the knowledge of the question type

- Explicit: adding the predicted question type
- Implicit: a multi-task learning approach (the model learns to predict the question type and the primary task of question summarization)

The proposed model outperforms T5, PEGASUS, and BART

› J Biomed Inform. 2022 Mar;127:104015. doi: 10.1016/j.jbi.2022.104015. Epub 2022 Feb 5.

## Predicting physician burnout using clinical activity logs: Model performance and lessons learned



Sunny S Lou <sup>1</sup>, Hanyang Liu <sup>2</sup>, Benjamin C Warner <sup>2</sup>, Derek Harford <sup>1</sup>, Chenyang Lu <sup>2</sup>, Thomas Kannampallil <sup>3</sup>

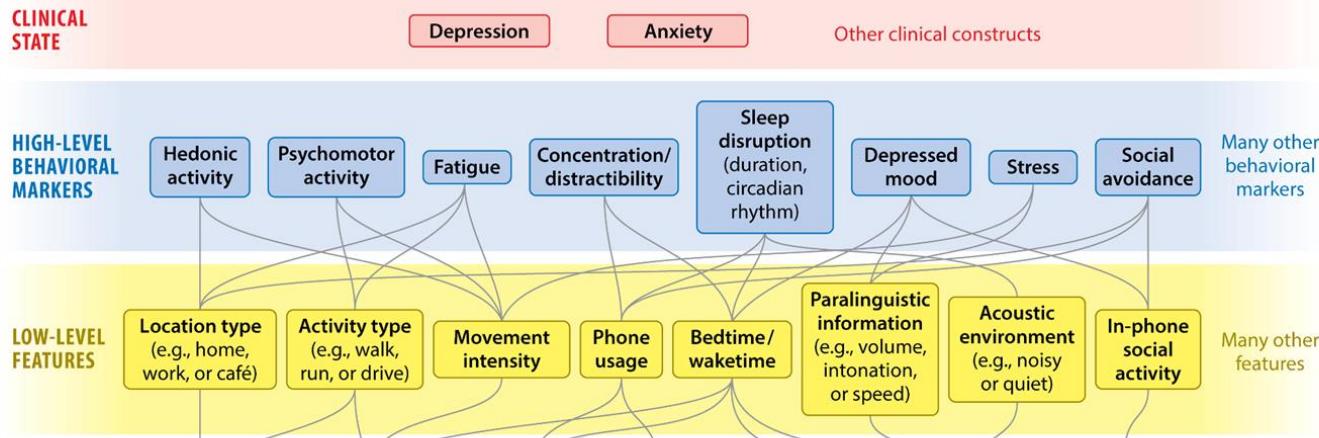
Affiliations + expand

PMID: 35134568 PMCID: PMC8901565 DOI: 10.1016/j.jbi.2022.104015



# Burnout

## Workload



Concentration,  
efficiency

### Workload features

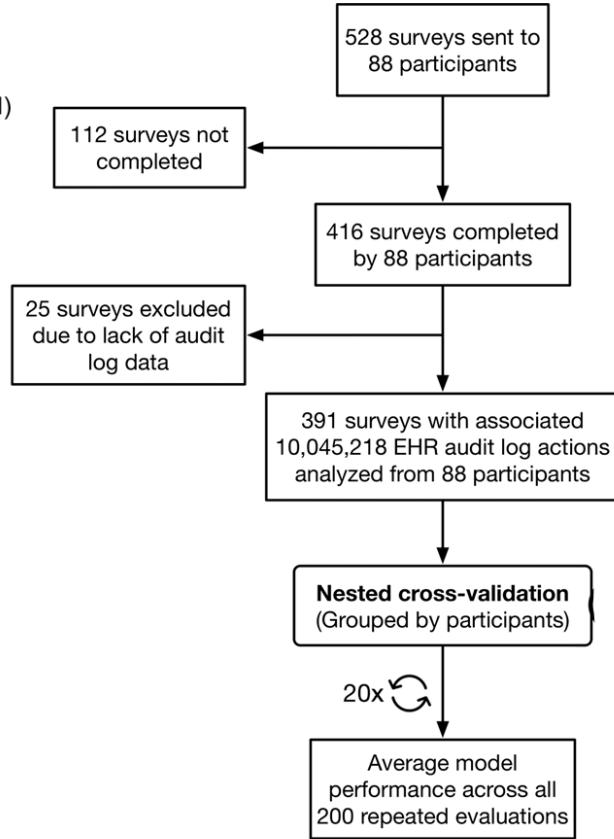
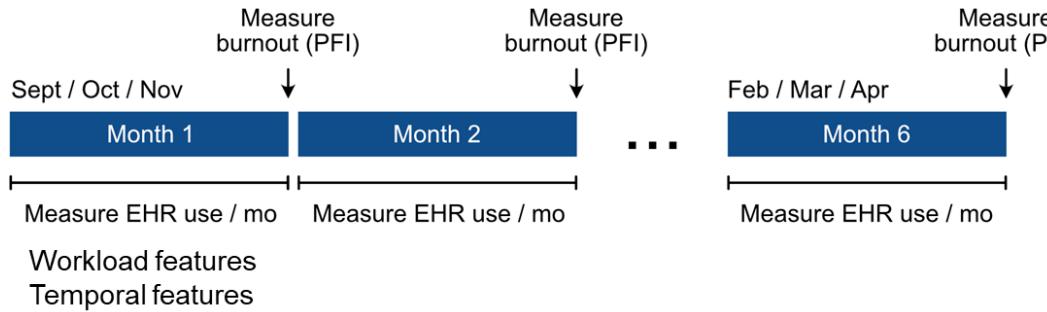
- Patient load
- Total EHR time
- After-hours time
- Inbox time
- Chart review time
- Note time
- Order volume

### Temporal features

Statistical features (i.e., mean, min, max, skew, kurtosis, energy, entropy, autocorrelation, slope):

- Pattern of EHR access events
- Pattern of switching between patient charts

Survey



Feature Set	Best Model	AUROC	Accuracy
Workload	Random Forest	0.595 (0.355, 0.808)	0.567 (0.393, 0.742)
Temporal	SVM	0.581 (0.343, 0.790)	0.556 (0.318, 0.756)
Workload + Temporal	GBM	0.583 (0.270, 0.831)	0.559 (0.386, 0.780)
1 <sup>st</sup> survey score	Neural Net	0.819 (0.551, 0.999)	0.765 (0.547, 0.952)
1 <sup>st</sup> survey score +workload	Neural Net	0.829 (0.607, 0.996)	0.781 (0.587, 0.936)

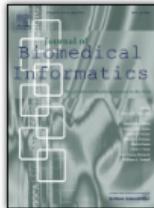
Address a significant public health concern

Develop machine learning-based digital phenotype of burnout using audit log files.

Machine learning models relying on EHR-based workload metrics had limited discriminative performance, highlighting the multi-factorial and highly-individualized nature of physician burnout

> J Biomed Inform. 2022 Apr;128:104039. doi: 10.1016/j.jbi.2022.104039. Epub 2022 Feb 26.

## Clinical notes: An untapped opportunity for improving risk prediction for hospitalization and emergency department visit during home health care



Jiyoun Song <sup>1</sup>, Mollie Hobensack <sup>2</sup>, Kathryn H Bowles <sup>3</sup>, Margaret V McDonald <sup>4</sup>, Kenrick Cato <sup>5</sup>, Sarah Collins Rossetti <sup>6</sup>, Sena Chae <sup>7</sup>, Erin Kennedy <sup>8</sup>, Yolanda Barrón <sup>9</sup>, Sridevi Sridharan <sup>10</sup>, Maxim Topaz <sup>11</sup>

Affiliations + expand

PMID: 35231649 PMCID: [PMC9825202](#) DOI: [10.1016/j.jbi.2022.104039](https://doi.org/10.1016/j.jbi.2022.104039)



# Release the hidden power of clinical notes

## Setting



Patients from the largest home health care (HHC) organization in New York between 2015 and 2017

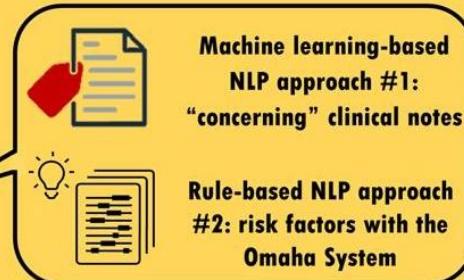
## Dataset



&



Structured dataset: OASIS and EHR  
Unstructured dataset: Clinical notes



## Risk modeling



- Risk Model 1: structured dataset
- Risk Model 2: structured dataset + clinical notes (NLP approach #1)
- Risk Model 3: structured dataset + clinical notes (NLP approach #2)
- Risk Model 4: structured dataset + clinical notes (NLP approach #1 and 2)



Each of the four risk models was trained using five machine learning algorithms (Logistic Regression, Random Forest, Bayesian network, SVM, and Naive Bayes)

## Findings



**Risk Model 4** shows an increase in the PRC area of 17.8% over **Risk Model 1**

In **Risk Model 3**, which incorporates the detailed risk factors of the Omaha System, the PRC area increased by 5% over **Risk Model 2**

Random Forest had the highest PRC area of 86.4% among the five machine learning algorithms

## Conclusion



Clinical notes combined with structured assessment data can improve predictive ability for hospitalization and ED visits in HHC

Acknowledgement: Agency for Healthcare Research and Quality [AHRQ] (R01 HS027742), "Building risk models for preventable hospitalizations and emergency department visits in homecare (Homecare-CONCERN)."

COLUMBIA UNIVERSITY  
IN THE CITY OF NEW YORK

VISITING NURSE SERVICE  
OF NEW YORK

Visiting Nurse Service  
of New York

# The power demonstration

## Result

- During the study period, 8373/86,823 (9.6%) HHC episodes resulted in hospitalization or ED visits.
- The SVM showed the highest F-score (0.82), while the Random Forest showed the highest PRC area (0.864).
- Adding information extracted from clinical notes significantly improved the risk prediction ability by up to 16.6% in F-score and 17.8% in PRC.

## Conclusion

- Information from clinical notes integrated with the structured data improved the ability to identify patients at risk for these emergent care events.

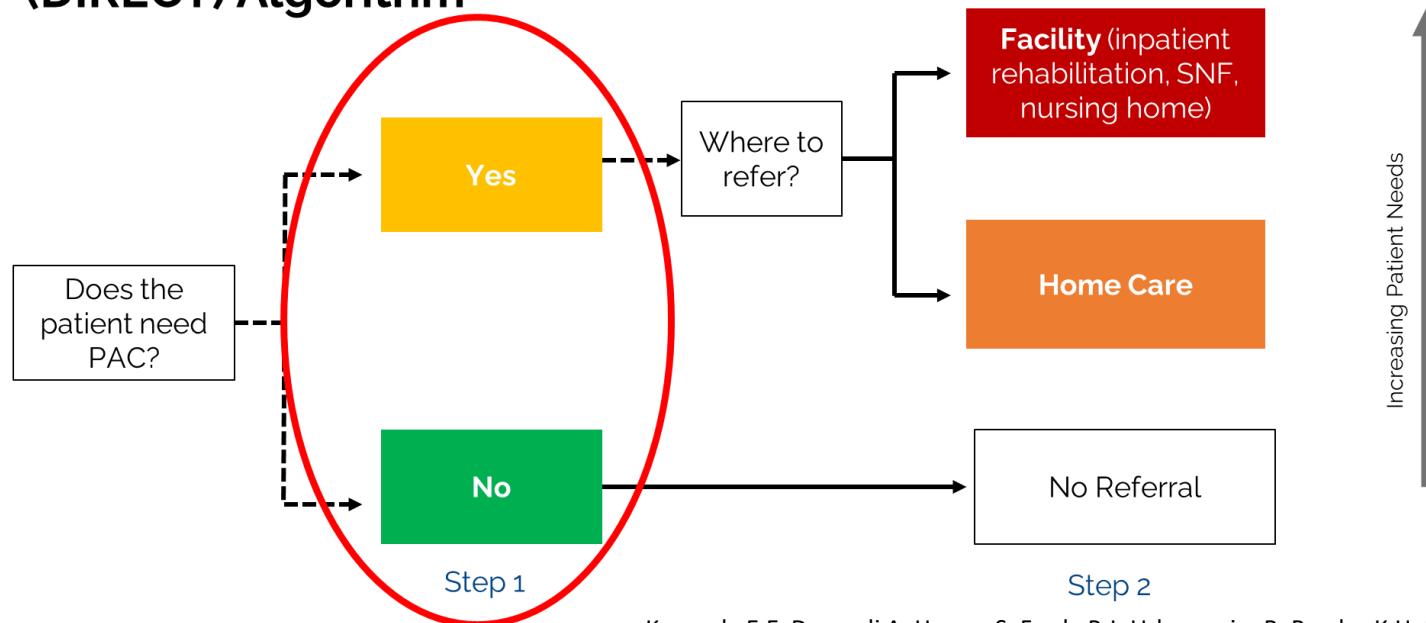
	Sensitivity (Precision)	PPV (Recall)	PRC Area
<b>Logistic regression</b>			
<i>Risk Model 1</i>	0.794	0.64	0.736
<i>Risk Model 2</i>	0.812	0.652	0.756
<i>Risk Model 3</i>	0.833	0.683	0.774
<i>Risk Model 4</i>	0.837	0.694	<b>0.812</b>
<b>Random Forest</b>			
<i>Risk Model 1</i>	0.896	0.692	0.818
<i>Risk Model 2</i>	0.909	0.693	0.84
<i>Risk Model 3</i>	0.918	0.707	0.845
<i>Risk Model 4</i>	0.927	0.721	<b>0.864</b>
<b>Bayes Network</b>			
<i>Risk Model 1</i>	0.721	0.643	0.71
<i>Risk Model 2</i>	0.749	0.708	0.757
<i>Risk Model 3</i>	0.815	0.72	0.795
<i>Risk Model 4</i>	0.827	0.762	<b>0.836</b>
<b>SVM</b>			
<i>Risk Model 1</i>	0.801	0.675	0.765
<i>Risk Model 2</i>	0.82	0.687	0.784
<i>Risk Model 3</i>	0.902	0.697	0.807
<i>Risk Model 4</i>	0.922	0.731	<b>0.821</b>
<b>Naïve Bayes</b>			
<i>Risk Model 1</i>	0.702	0.65	0.688
<i>Risk Model 2</i>	0.721	0.677	<b>0.701</b>
<i>Risk Model 3</i>	0.692	0.661	0.682
<i>Risk Model 4</i>	0.702	0.682	0.684

Red arrows and percentage increases:

- +10.3% (from Risk Model 1 to Risk Model 4) for Logistic regression
- +5.6% (from Risk Model 1 to Risk Model 4) for Random Forest
- +17.8% (from Risk Model 1 to Risk Model 4) for Bayes Network
- +7.3% (from Risk Model 1 to Risk Model 4) for SVM
- +1.9% (from Risk Model 1 to Risk Model 4) for Naïve Bayes

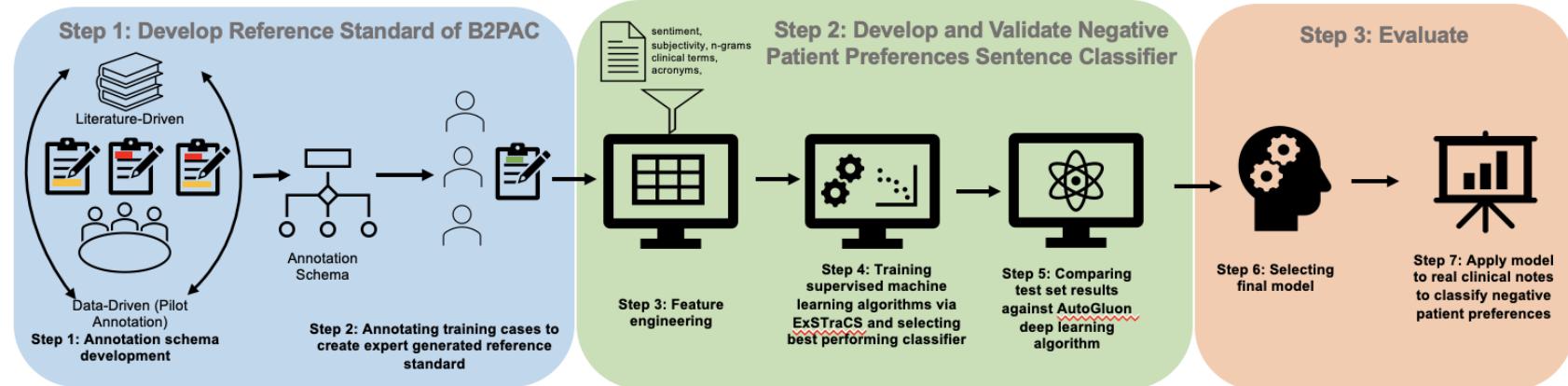
# Nursing Informatics

## Discharge Referral Expert System for Care Transitions (DIRECT) Algorithm



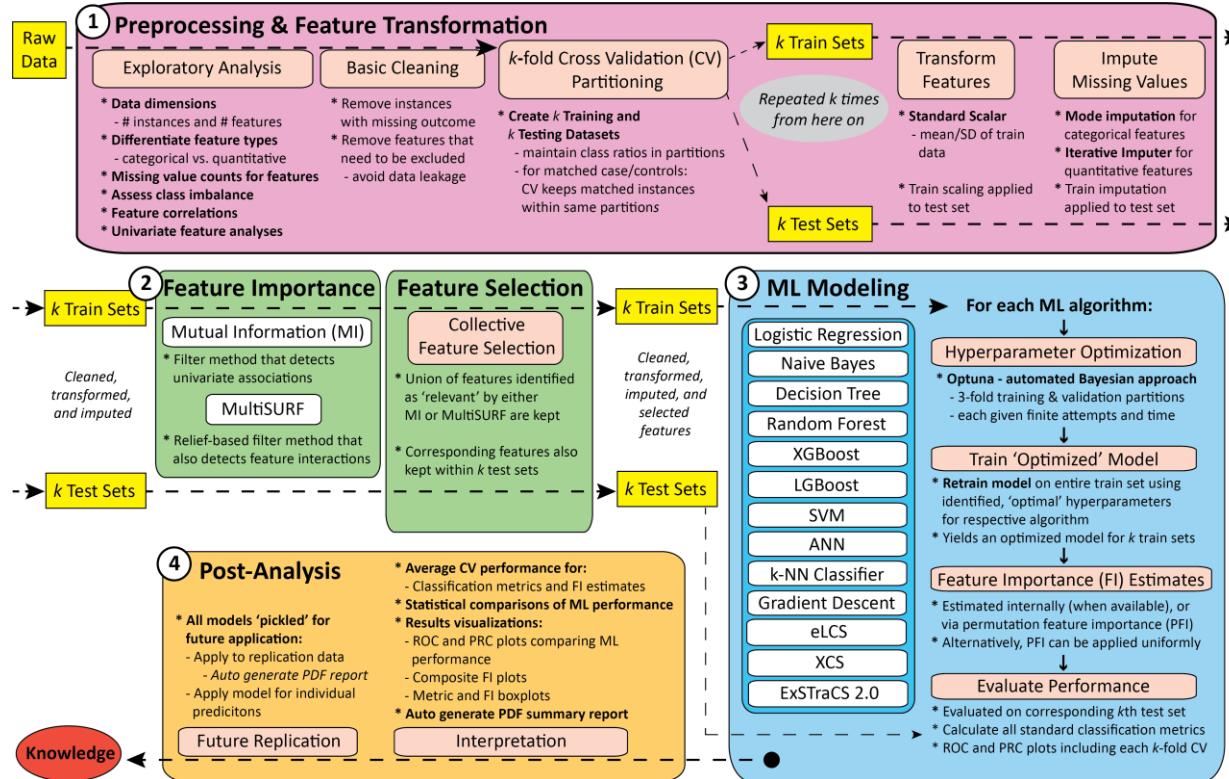
Kennedy E.E, Davoudi A, Hwang S, Freda P.J, Urbanowicz R, Bowles K.H, Mowery D.L. Identifying Barriers to Post-Acute Care Referral and Characterizing Negative Patient Preferences Among Hospitalized Older Adults Using Natural Language Processing. AMIA 2022

# Process Workflow for Study Methods



Kennedy E.E, Davoudi A, Hwang S, Freda P.J, Urbanowicz R, Bowles K.H, Mowery D.L. Identifying Barriers to Post-Acute Care Referral and Characterizing Negative Patient Preferences Among Hospitalized Older Adults Using Natural Language Processing. AMIA 2022

# Supervised Learning Pipeline Using AutoML-BC



› JAMIA Open. 2022 May 26;5(2):ooac034. doi: 10.1093/jamiaopen/ooac034. eCollection 2022 Jul.

## Do nurses document all discussions of patient problems and nursing interventions in the electronic health record? A pilot study in home healthcare

Jiyoun Song <sup>1</sup>, Maryam Zolnoori <sup>1</sup>, Danielle Scharp <sup>1</sup>, Sasha Vergez <sup>2</sup>, Margaret V McDonald <sup>1,2</sup>,  
Sridevi Sridharan <sup>2</sup>, Zoran Kostic <sup>3</sup>, Maxim Topaz <sup>1,2,4</sup>

Affiliations + expand

PMID: 35663115 PMCID: PMC9154272 DOI: 10.1093/jamiaopen/ooac034

# Nurse documentation of patient problem discussions and nursing interventions in the EHR



- To assess what health problems or nursing interventions were discussed during home health care (HHC) visits
- To examine the information overlap between patient-nurse verbal communication and EHR documentation (i.e., clinical notes and structured data)

Speaker	Time stamp	Conversation	Omaha system Problem	Omaha system Intervention
Patient	S2 16:40	I was to take the-- I was to take the hydrochlorothiazide. But I'll tell you I fell asleep. <b>I didn't take anything.</b>	Medication regimen: does not follow recommended dosage/schedule	
Nurse	S1 16:48	Yeah, take it because otherwise, <b>the leg will swell up more.</b>	Circulation: edema	Teaching, Guidance, and Counseling: medication action/side effects

Song J, Zolnoori M, Scharp D, et al. Do nurses document all discussions of patient problems and nursing interventions in the electronic health record? A pilot study in home healthcare. JAMIA Open. 2022;5(2)

# Nurse documentation of patient problem discussions and nursing interventions in the EHR



- Example #1. Identify problem through verbal communication was **accordingly documented** in the narrative notes and/or structured data
- Example #2. Identify problem through verbal communication was **completely missed** in both narrative notes and structured data

Speaker	Time stamp	Verbal communication	Narrative Note	Structured data
		Verbal communication with Omaha system annotation		
Nurse	S1 01:34	Show me how you walk. You feel now <ul style="list-style-type: none"><li>▪ Omaha system Intervention (<b>Surveillance: gait training</b>)</li></ul>		Functional Indicate Musculoskeletal Assessment: (Mark All That Apply) <ul style="list-style-type: none"><li><input type="radio"/> Decreased Strength in What Extremities</li></ul>
Patient	S2 09:04	I can't get in a car. I can't move. <ul style="list-style-type: none"><li>▪ Omaha system Problem (<b>Neuro-musculoskeletal function: decreased muscle strength, decreased sensation</b>)</li></ul>	Pt ambulates with a walker. Not steady on both feet yet.	Does Decreased Strength Exist? (Mark All That Apply) <ul style="list-style-type: none"><li><input type="radio"/> Lower Bilateral Extremities</li></ul>
Patient	S2 12:08	I can walk, but I need a walker. <ul style="list-style-type: none"><li>▪ Omaha system Problem (<b>Neuro-musculoskeletal function: decreased muscle strength</b>)</li></ul>		
Nurse	S1 12:09	You cannot walk safely. <ul style="list-style-type: none"><li>▪ Omaha system Problem (<b>Neuro-musculoskeletal function: decreased muscle strength</b>)</li></ul>		

Speaker	Time stamp	Verbal communication	Narrative Note	Structured data
		Verbal communication with Omaha system annotation		
Nurse	S1 00:34	Your son told me that your ear feels funny. You said something? <ul style="list-style-type: none"><li>▪ Omaha system Intervention (<b>Surveillance: signs/symptoms-physical</b>)</li></ul>		Were head and neck assessed? <ul style="list-style-type: none"><li><input type="radio"/> Yes</li></ul>
Patient	S2 00:39	Little muffled. <ul style="list-style-type: none"><li>▪ Omaha system Problem (<b>Hearing: Other</b>)</li></ul>		Indicate head and neck assessment findings: (Mark All That Apply) <ul style="list-style-type: none"><li><input type="radio"/> No Problems Identified Eyes/Ears/Nose/Throat</li></ul>
Nurse	S1 00:39	Yeah. Like the sound, you mean? <ul style="list-style-type: none"><li>▪ Omaha system Intervention (<b>Surveillance: signs/symptoms-physical</b>)</li></ul>		
Patient	S2 00:42	Yeah. This side now. <ul style="list-style-type: none"><li>▪ Omaha system Problem (<b>Hearing: Other</b>)</li></ul>	None	

# Nurse documentation of patient problem discussions and nursing interventions in the EHR



- **Result**

- Average 10 different problems (range 4 – 23) and 22.4 different nursing interventions (range 8 – 49) were identified during verbal communications
- The most frequent problems mentioned were “Circulation,” “Skin,” and “Medication regimen.”
- The most common interventions discussed were “Surveillance,” followed by “Teaching, Guidance, and Counseling”
- Approximately **50.5% of the problems and 20.8% of the intervention mentioned verbally were not documented in the EHR.**

	Verbal communication	Clinical notes	Structured EHR data	All EHR (Clinical notes and/or structured data)	Discussed verbally but not documented in EHR*
# of problems	216	64 (29.6%)	91 (42.1%)	120 (55.6%)	<b>50.5%</b> (109/216)
# of interventions	490	291 (59.4%)	222 (45.3%)	390 (79.6%)	<b>20.8%</b> (102/490)
Total	706	355 (50.3%)	313 (44.3%)	510 (72.2%)	<b>29.9%</b> (211/406)

- **Conclusion**

- Our results highlight the value of analyzing communications between HHC patients and nurses. Future studies should explore ways to capture information in verbal communication using automated speech recognition.

# Imaging Informatics

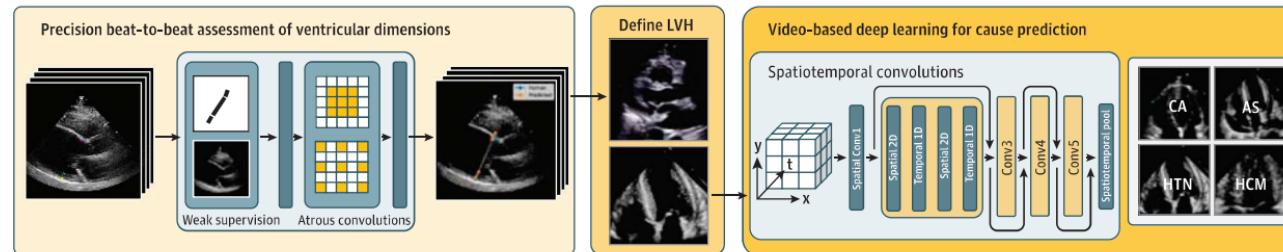
# AI Enables Precision Disease Detection at Scale

JAMA Cardiology | Original Investigation

## High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy With Cardiovascular Deep Learning

Figure 1. Deep Learning Workflow Combining Evaluation of Ventricular Dimensions and Suspicion for Underdiagnosed Diseases

A Deep learning workflow

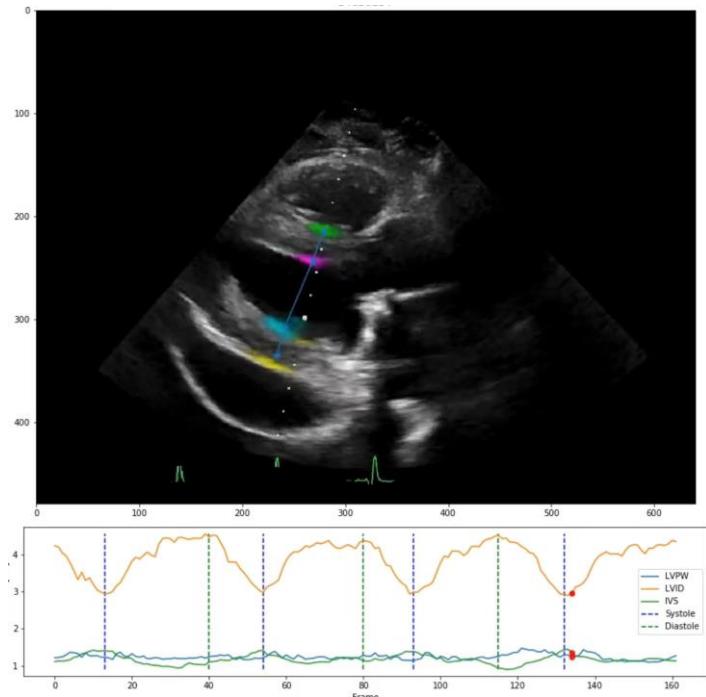
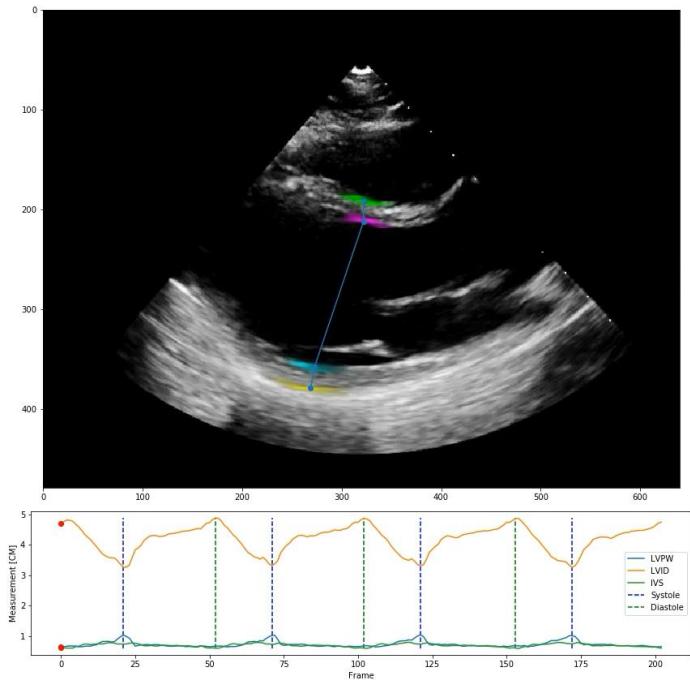


### Precision Identification of Hypertrophy

### AI Texture Based Evaluation for Cardiac Amyloidosis

Duffy\*, Cheng\*, Yuan\* et al. **JAMA Cardiology**. (Feb 23, 2022)

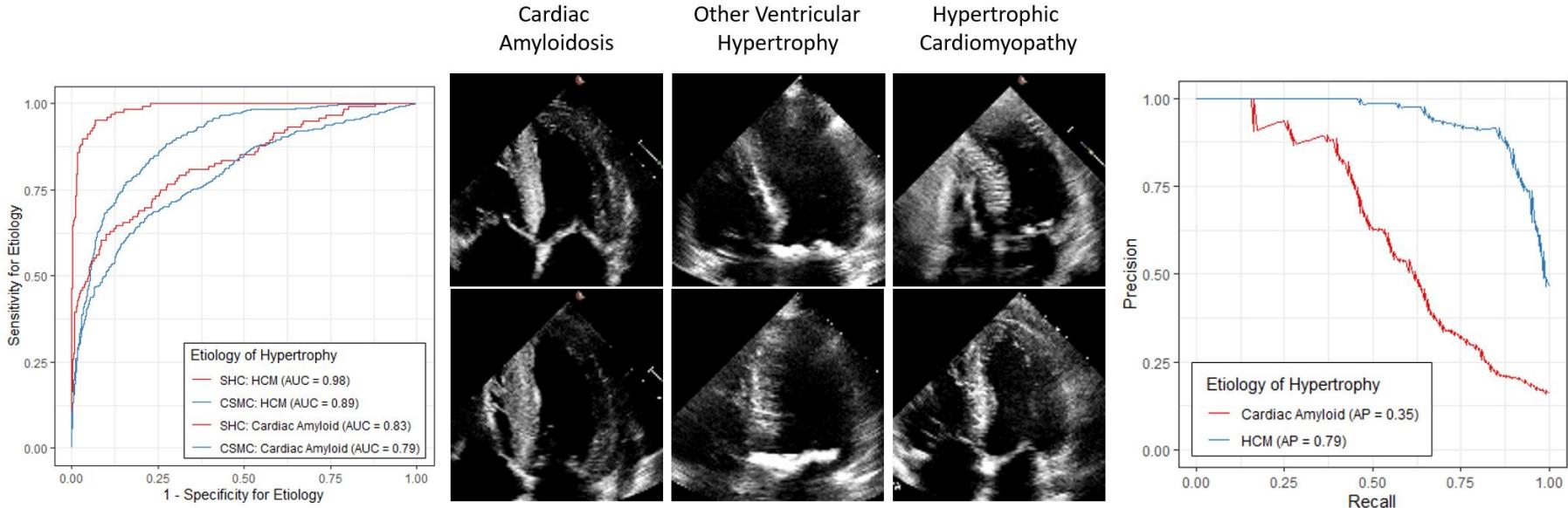
# AI Precision Assessment of Hypertrophy



<https://arxiv.org/abs/2106.12511>

Duffy\*, Cheng\*, Yuan\* et al. **JAMA Cardiology**. (Feb 23, 2022)

# Augmenting Clinical Intuition

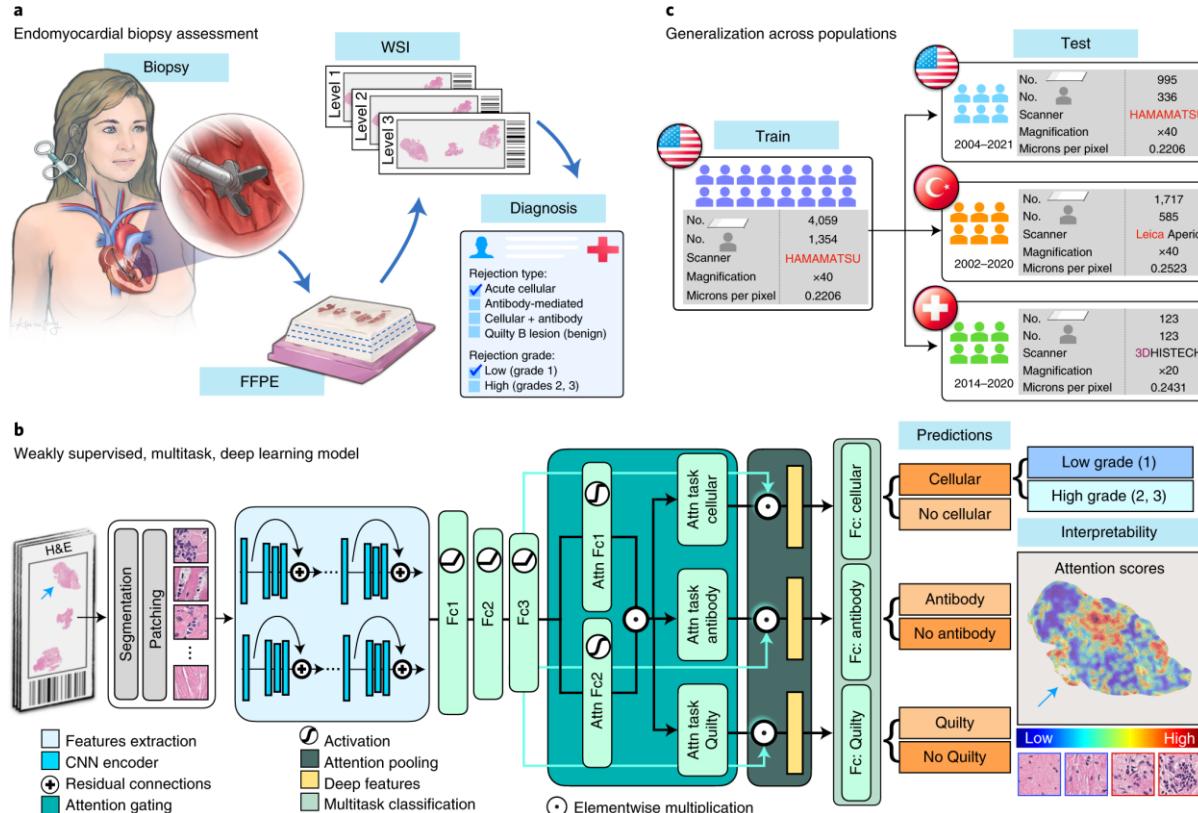


> Nat Med. 2022 Mar;28(3):575-582. doi: 10.1038/s41591-022-01709-2. Epub 2022 Mar 21.

## Deep learning-enabled assessment of cardiac allograft rejection from endomyocardial biopsies

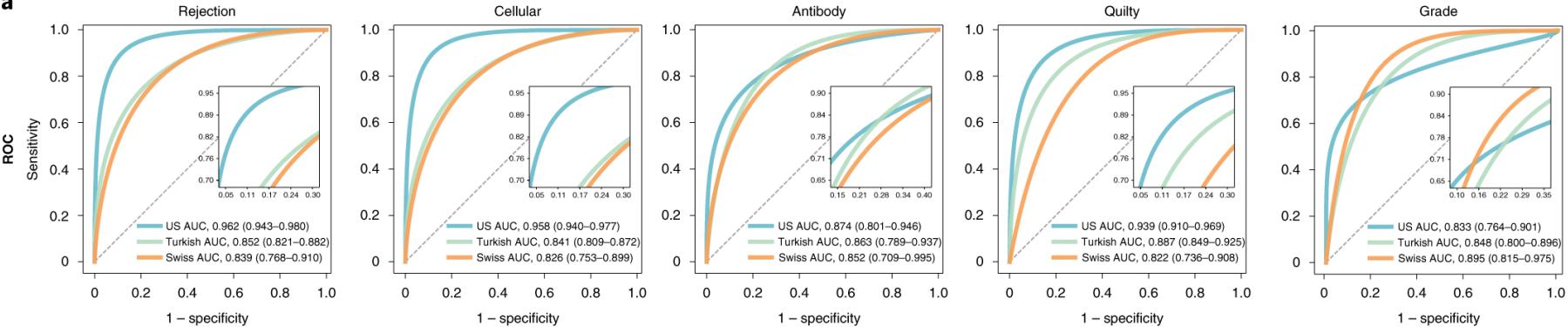
Jana Lipkova # 1 2 3, Tiffany Y Chen # 1 2 3, Ming Y Lu 1 2 3 4, Richard J Chen 1 2 3 5,  
Maha Shady 1 2 3 5, Mane Williams 1 2 3 5, Jingwen Wang 1 6, Zahra Noor 1,  
Richard N Mitchell 1 7, Mehmet Turan 8, Gulfize Coskun 8, Funda Yilmaz 9, Derya Demir 9,  
Deniz Nart 9, Kayhan Basak 10, Nesrin Turhan 10, Selvinaz Ozkara 10, Yara Banz 11,  
Katja E Odening 12 13, Faisal Mahmood 14 15 16 17 18

# Cardiac Rejection Assessment Neural Estimator (CRANE)

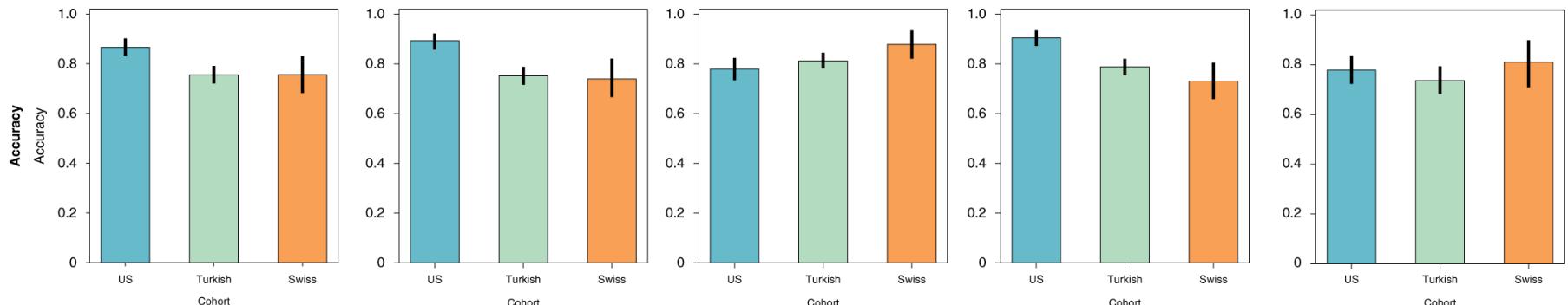


# Performance of the CRANE model at patient level

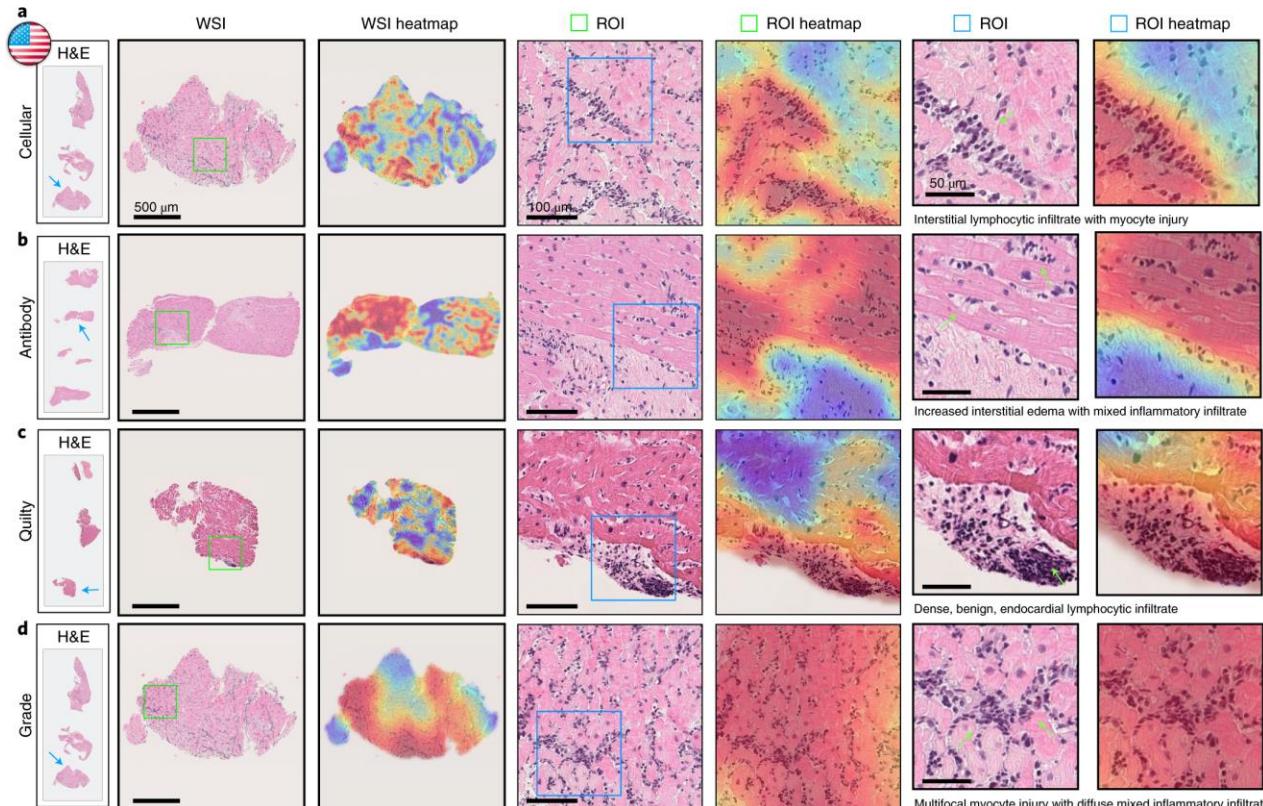
a



b



# Visualization of the attention heatmaps for the US cohort



Observational Study > Lancet Digit Health. 2022 Jan;4(1):e8-e17.

doi: 10.1016/S2589-7500(21)00215-6.

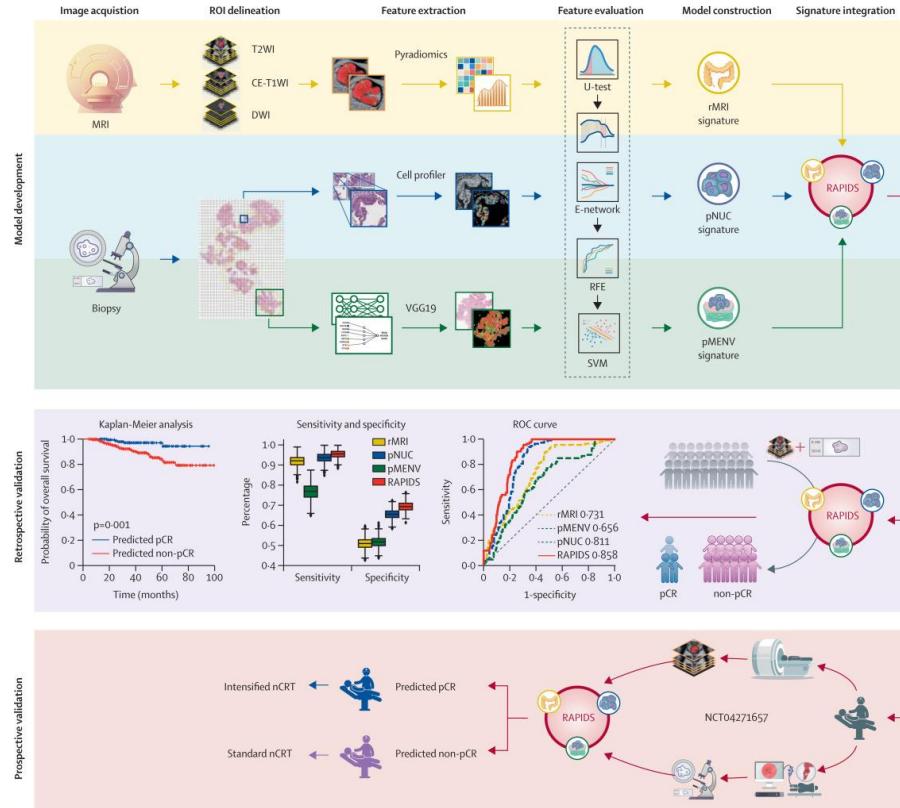
## Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study

Lili Feng <sup>1</sup>, Zhenyu Liu <sup>2</sup>, Chaofeng Li <sup>3</sup>, Zhenhui Li <sup>4</sup>, Xiaoying Lou <sup>5</sup>, Lizhi Shao <sup>6</sup>,  
Yunlong Wang <sup>1</sup>, Yan Huang <sup>5</sup>, Haiyang Chen <sup>7</sup>, Xiaolin Pang <sup>7</sup>, Shuai Liu <sup>7</sup>, Fang He <sup>7</sup>,  
Jian Zheng <sup>7</sup>, Xiaochun Meng <sup>8</sup>, Peiyi Xie <sup>8</sup>, Guanyu Yang <sup>9</sup>, Yi Ding <sup>10</sup>, Mingbiao Wei <sup>1</sup>,  
Jingping Yun <sup>11</sup>, Mien-Chie Hung <sup>12</sup>, Weihua Zhou <sup>13</sup>, Daniel R Wahl <sup>14</sup>, Ping Lan <sup>15</sup>, Jie Tian <sup>16</sup>,  
Xiangbo Wan <sup>17</sup>

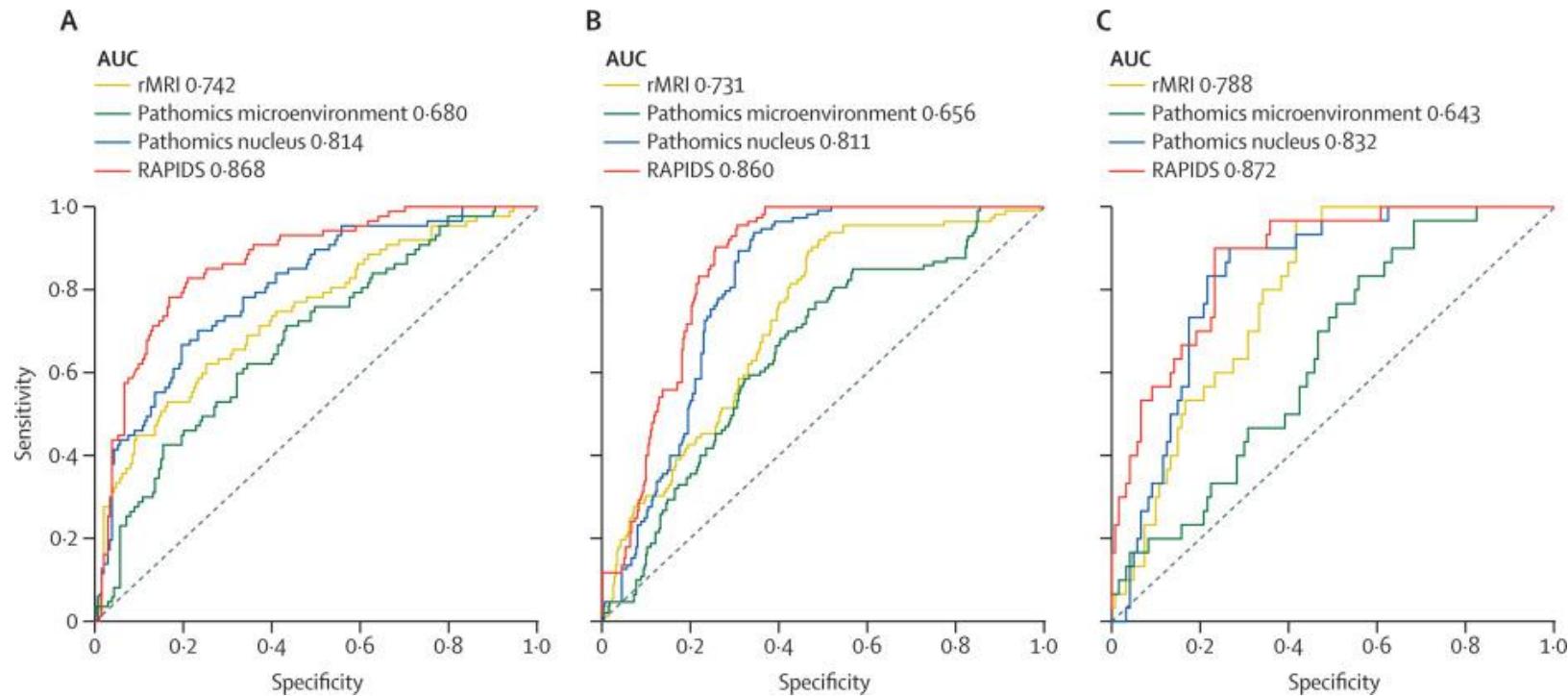
Affiliations + expand

PMID: 34952679 DOI: 10.1016/S2589-7500(21)00215-6

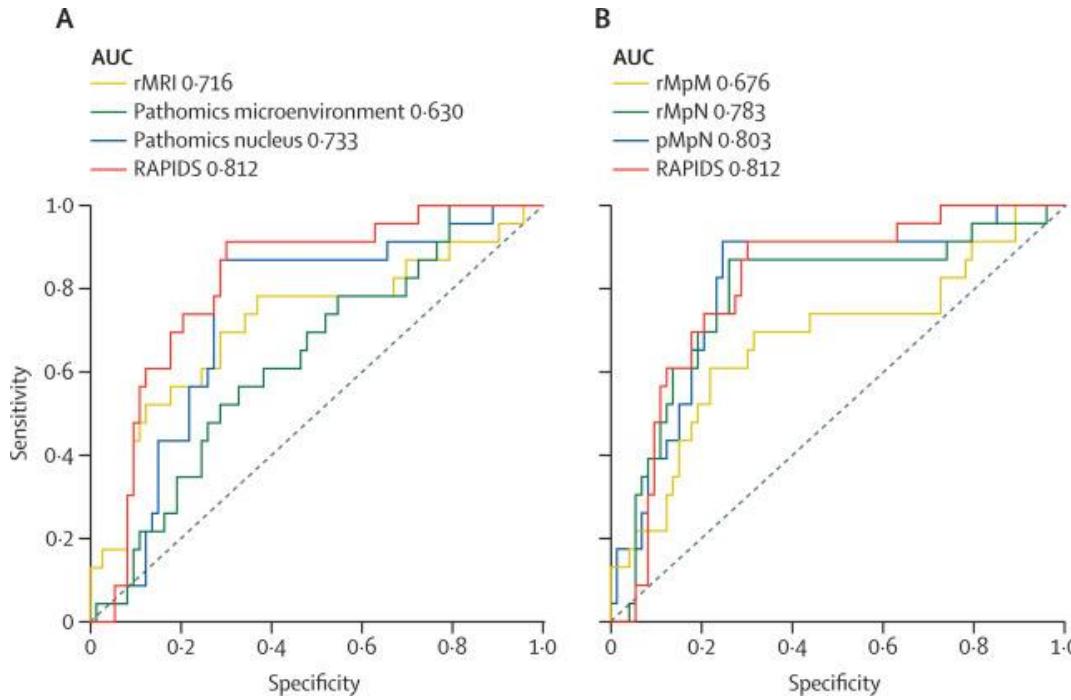
# Radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy



# Retrospective training and validation cohorts



# Prospective observational study



# Natural Language Processing

# Natural Language Processing



› NPJ Digit Med. 2022 Dec 26;5(1):194. doi: 10.1038/s41746-022-00742-2.

## A large language model for electronic health records

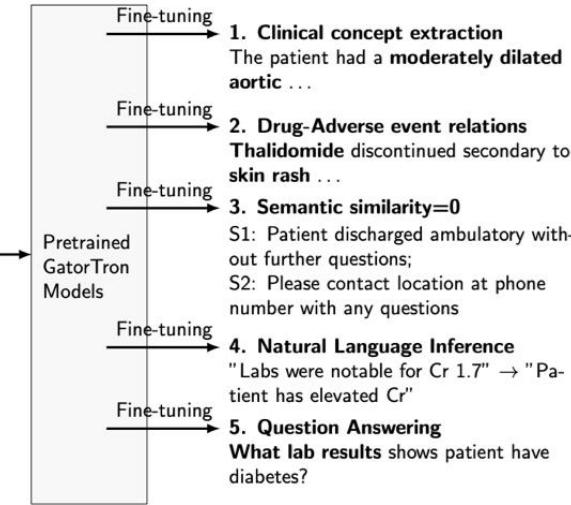
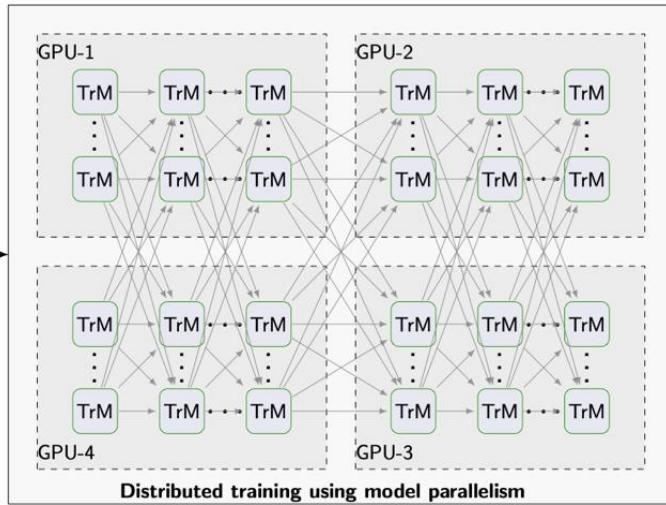


Xi Yang <sup>1 2</sup>, Aokun Chen <sup>1 2</sup>, Nima PourNejatian <sup>3</sup>, Hoo Chang Shin <sup>3</sup>, Kaleb E Smith <sup>3</sup>, Christopher Parisien <sup>3</sup>, Colin Compas <sup>3</sup>, Cheryl Martin <sup>3</sup>, Anthony B Costa <sup>3</sup>, Mona G Flores <sup>3</sup>, Ying Zhang <sup>4</sup>, Tanja Magoc <sup>5</sup>, Christopher A Harle <sup>1 5</sup>, Gloria Lipori <sup>5 6</sup>, Duane A Mitchell <sup>6</sup>, William R Hogan <sup>1</sup>, Elizabeth A Shenkman <sup>1</sup>, Jiang Bian <sup>1 2</sup>, Yonghui Wu <sup>7 8</sup>

Affiliations + expand

PMID: 36572766 PMCID: PMC9792464 DOI: 10.1038/s41746-022-00742-2

82 Billion clinical words  
6 Billion PubMed words  
2.5 Billion Wiki words



# Training and validation loss (GatorTron(s))

Training GatorTron-large model required ~6 days on 992 GPUs (A100 80Gb) from 124 NVIDIA SuperPOD nodes reference cluster architecture. !!!!!

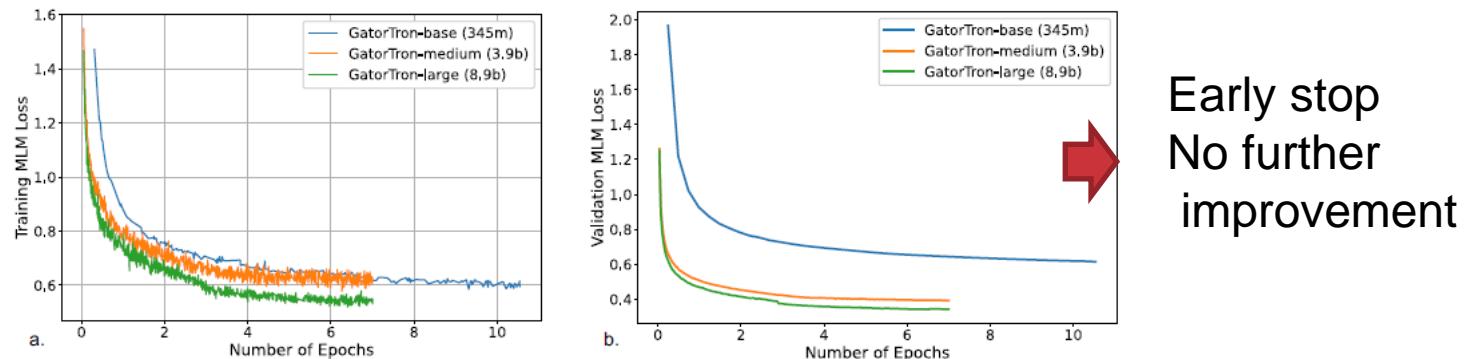


Fig. 2 Training loss and validation loss for GatorTron-base (345 million), medium (3.9 billion), and large (8.9 billion) models. a Training loss. b Validation loss. MLM masked language modeling.

The GatorTron-base model converged in 10 epochs, whereas the medium and large models converged in 7 epochs, faster convergence for larger models

# GATORTRON VS OTHERS (METRICS)

**Table 1.** Comparison of GatorTron with existing biomedical and clinical transformer models for clinical concept extraction and medical relation extraction.

Transformer	Clinical concept extraction									Medical relation extraction		
	2010 i2b2 <sup>39</sup>			2012 i2b2 <sup>40</sup>			2018 n2c2 <sup>41</sup>			2018 n2c2 <sup>41</sup>		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
BioBERT	0.8693	0.8653	0.8673	0.7478	0.8037	0.7747	0.8634	0.8921	0.8775	0.9663	0.9451	0.9555
ClinicalBERT	NA	NA	0.8780	NA	NA	0.7890	0.8592	0.8832	0.8710	0.9678	0.9414	0.9544
BioMegatron	0.8614	0.8761	0.8687	0.7591	0.8031	0.7805	0.8707	0.8915	0.8810	0.9711	0.9434	0.9571
GatorTron-base (1/4 data)	0.8682	0.9046	0.8860	0.7514	0.8013	0.7755	0.8772	0.8992	0.8881	0.9724	0.9457	0.9589
GatorTron-base	0.8748	0.9043	0.8893	0.7644	0.8221	0.7922	0.8759	0.9038	0.8896	0.9719	0.9482	0.9599
GatorTron-medium	0.8869	0.9122	0.8994	0.7812	0.8245	0.8022	0.8954	0.9035	0.8994	0.9721	0.9503	0.9611
GatorTron-large	0.8880	0.9116	<b>0.8996</b>	0.7862	0.8333	<b>0.8091</b>	0.8979	0.9021	<b>0.9000</b>	0.9776	0.9482	<b>0.9627</b>

Clinical concepts in 2010 i2b2 and 2012 i2b2 challenges: problems, treatments, lab tests; clinical concepts in 2018 n2c2 challenge: drugs, adverse events, and drug-related attributes (e.g., dose). Medical relation in 2018 n2c2 challenge: drug induced adverse events. Best F1 scores are presented in bold. NA: scores not reported.

GatorTron models outperformed existing biomedical and clinical transformer models performance

# Useful?



## Findings

- Bigger is better - Scale up the size of training data and model size improved performance.
- One model architecture solves 5 different clinical NLP tasks with state-of-the-art performance.
- Larger language models help more in complex NLP tasks such as question answering.

## Challenges

- Security and privacy of PHI data
- High-performance computing support

## Potentials to healthcare

- Patient information extraction
- Assist in documenting patient reports, and reduce documentation burden
- Assist in generating treatment plans.
- Next-generation intelligent EHR systems with conversational ability.

› J Am Med Inform Assoc. 2022 Jun 14;29(7):1208-1216. doi: 10.1093/jamia/ocac040.

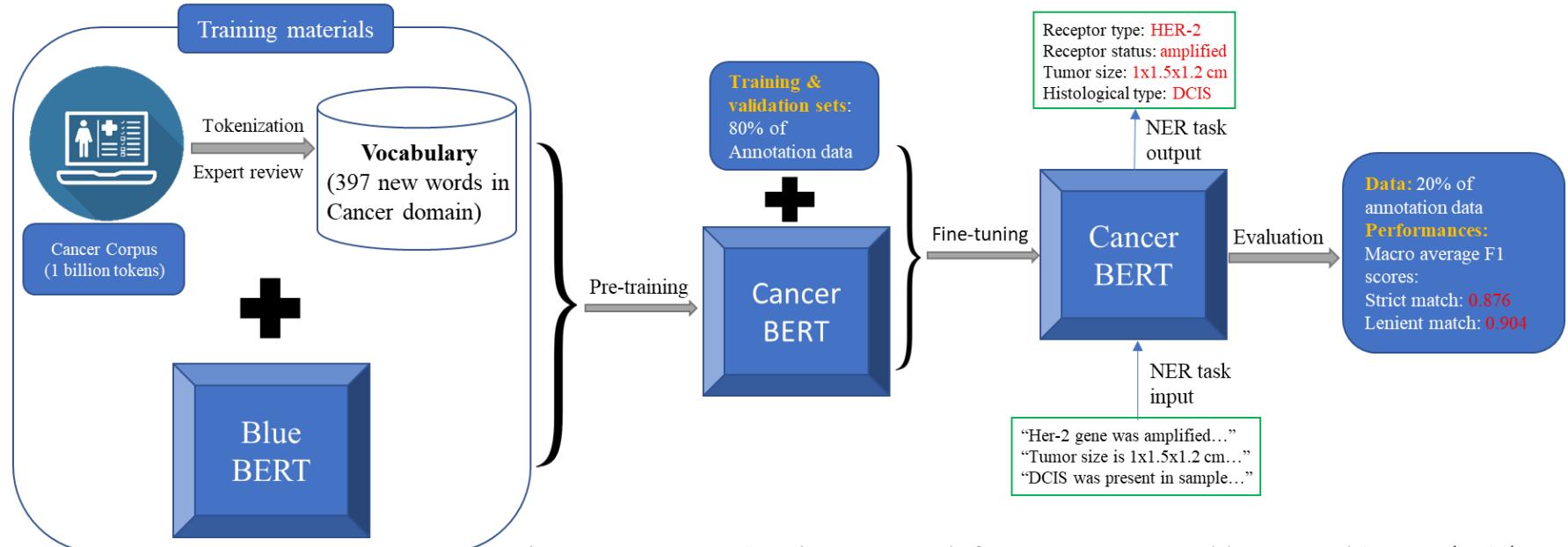
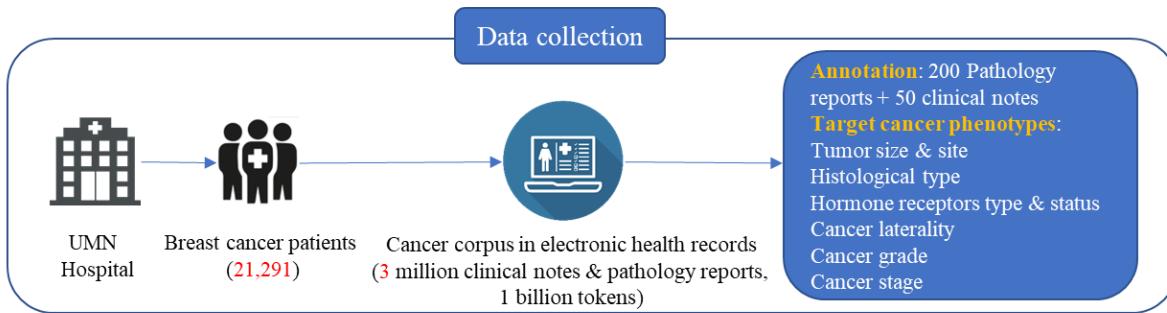
## CancerBERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records



Sicheng Zhou <sup>1</sup>, Nan Wang <sup>2</sup>, Liwei Wang <sup>3</sup>, Hongfang Liu <sup>3</sup>, Rui Zhang <sup>1</sup> <sup>4</sup>

Affiliations + expand

PMID: 35333345 PMCID: [PMC9196678](#) DOI: [10.1093/jamia/ocac040](https://doi.org/10.1093/jamia/ocac040)



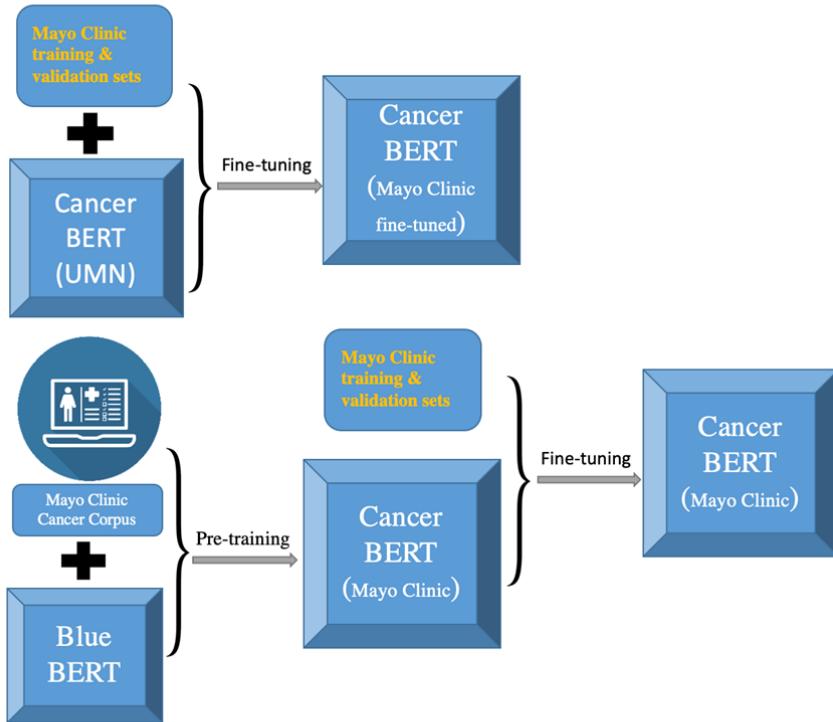
Zhou S, Wang N, Wang L, Liu H, Zhang R. J Am Med Inform Assoc. 2022 Jun 14;29(7):1208-1216. doi: 10.1093/jamia/ocac040.

# Results-Model performance

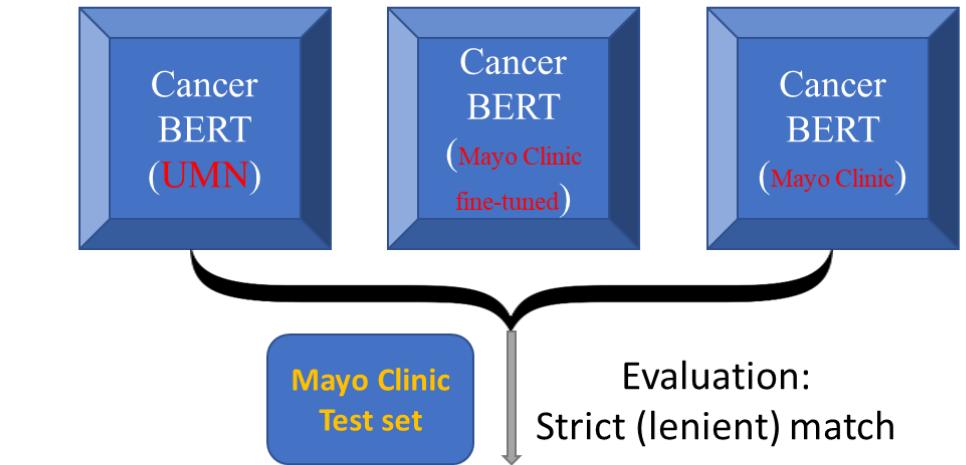
Entity type	BiLSTM-CRF	BERT-large Origin	BlueBERT (PubMed+MIMIC III)	BioBERT (PubMed)	character-BERT (Medical)	CancerBERT <sub>OrigVoc</sub> (EHRs corpus)	CancerBERT <sub>CustVoc_997</sub> (EHRs corpus)	CancerBERT <sub>CustVoc_397</sub> (EHRs corpus)
<b>Hormone</b>	0.953 (0.957)	0.976 (0.985)	0.979 (0.984)	0.982 (0.987)	0.972 (0.983)	<b>0.984 (0.988)</b>	0.979 (0.985)	0.982 (0.985)
<b>Receptor type</b>								
<b>Hormone</b>	0.856 (0.856)	0.846 (0.846)	0.885 (0.885)	0.859 (0.859)	0.851 (0.851)	<b>0.901* (0.901*)</b>	0.887 (0.887)	0.891 (0.891)
<b>Receptor status</b>								
<b>Tumor size</b>	0.664 (0.709)	0.663 (0.767)	0.781 (0.819)	<b>0.785</b> (0.821)	0.674 (0.684)	0.765 (0.813)	0.784 (0.824)	0.781 ( <b>0.827*</b> )
<b>Tumor site</b>	0.562 (0.771)	0.696 (0.769)	0.711 (0.797)	<b>0.749*</b> (0.799)	0.688 (0.762)	0.733 (0.792)	0.715 (0.787)	0.727 ( <b>0.824*</b> )
<b>Tumor grade</b>	0.910 (0.910)	0.857 (0.857)	0.891 (0.891)	0.886 (0.886)	0.833 (0.833)	0.891 (0.891)	0.898 (0.898)	<b>0.915* (0.915*)</b>
<b>Tumor laterality</b>	0.935 (0.935)	0.926 (0.926)	0.931 (0.931)	0.943 (0.943)	0.934 (0.934)	0.939 (0.939)	0.947 (0.947)	<b>0.953* (0.953*)</b>
<b>Cancer stage</b>	0.908 (0.908)	0.804 (0.804)	0.870 (0.870)	0.869 (0.869)	<b>0.907 (0.907)</b>	0.870 (0.870)	0.885 (0.885)	0.898 (0.898)
<b>Histological type</b>	<b>0.885*</b> (0.938)	0.823 (0.918)	0.843 (0.922)	0.855 (0.934)	0.861 ( <b>0.943*</b> )	0.849 (0.922)	0.862 (0.937)	0.862 (0.938)
<b>Macro average</b>	0.834 (0.873)	0.824 (0.859)	0.862 (0.887)	0.868 (0.889)	0.840 (0.862)	0.867 (0.889)	0.871 (0.896)	<b>0.876* (0.904*)</b>
<b>Micro average</b>	0.876 (0.905)	0.873 (0.907)	0.898 (0.921)	0.904 (0.926)	0.883 (0.906)	0.903 (0.925)	0.906 (0.930)	<b>0.909* (0.933*)</b>

Note: The scores were averaged scores based on 10 runs, \* indicates statistically higher than other methods (CI: 0.95).

## Model development at Mayo Clinic



## Model Evaluations



	CancerBERT (UMN)	CancerBERT (Mayo Clinic fine-tuned)	CancerBERT (Mayo Clinic)
Macro-average F1	0.828 (0.849)	0.892 (0.930)	0.898* (0.932)
Micro-average F1	0.864 (0.906)	0.925 (0.947)	0.932* (0.952)

CancerBERT models trained on a single institute can be transferred to another institute and achieved comparable performance compared to locally trained models.

› JAMA Netw Open. 2022 Jan 4;5(1):e2144967. doi: 10.1001/jamanetworkopen.2021.44967.

## Examination of Stigmatizing Language in the Electronic Health Record



Gracie Himmelstein <sup>1</sup> <sup>2</sup>, David Bates <sup>3</sup> <sup>4</sup>, Li Zhou <sup>3</sup>

Affiliations + expand

PMID: 35084481 PMCID: [PMC8796019](#) DOI: [10.1001/jamanetworkopen.2021.44967](#)

# Study setting

---

## Why does stigmatizing language in the medical record matter?

- “It is in talking and writing to other doctors about patients that cultural assumptions, beliefs, and values are most directly displayed.” (Aspach, 1988)
- These assumptions and beliefs are transmitted to subsequent providers through the use of stigmatizing language

**Research Questions:** What is the prevalence of stigmatizing language in hospital admissions notes? Are specific demographic groups more likely to have stigmatizing language in their medical records? Does this vary by medical condition?

**Data:** Free text of 48,651 admission notes for 29,783 unique patients written by 1,932 clinicians at a large, urban, academic medical center in 2018.

# Methods and Results



## Methods

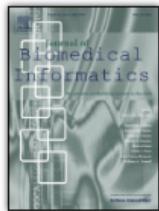
- Assembled lists of stigmatizing language from published sources
- Assessed the association between stigmatizing language and patient/provider characteristics

## Results

- Stigmatizing language appeared in 2.5% of all notes – this varied by patient's medical conditions 6.9% among patients with diabetes, 3.4% among patients with substance use, and 0.9% among patients with chronic pain
- Overall notes about Black (vs. White) patients had a 0.67 percentage point greater probability of containing stigmatizing language, with similar disparities in all three diagnosis-specific subgroups.
- Among physicians, more years of training was associated with less use of stigmatizing language overall (-0.05 percentage points/year)

› J Biomed Inform. 2022 Mar;127:103984. doi: 10.1016/j.jbi.2021.103984. Epub 2022 Jan 7.

## Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing



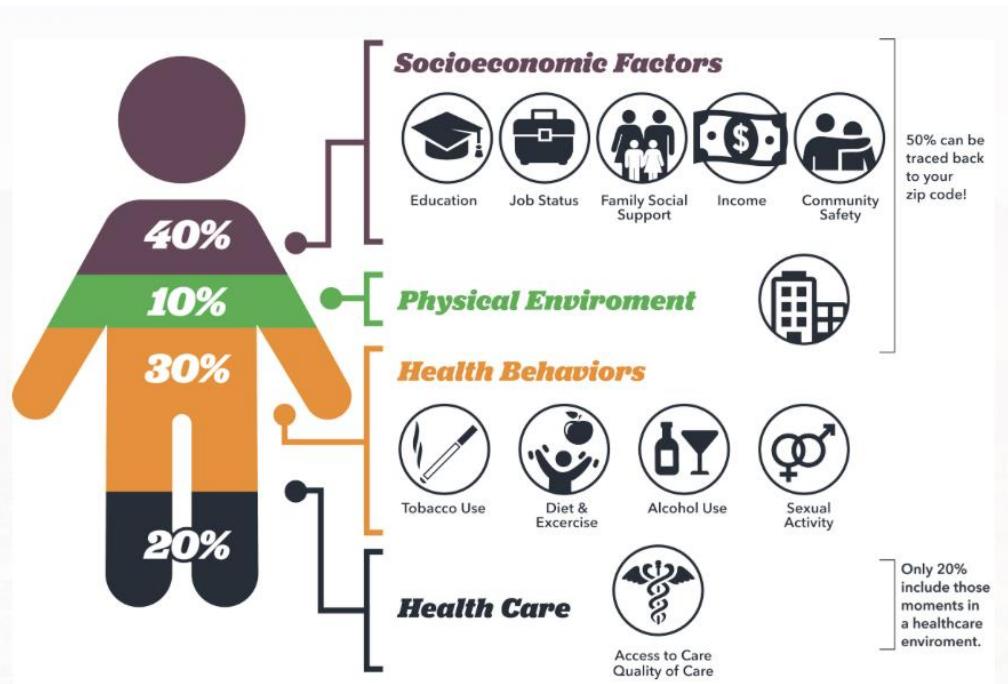
Sifei Han <sup>1</sup>, Robert F Zhang <sup>2</sup>, Lingyun Shi <sup>1</sup>, Russell Richie <sup>1</sup>, Haixia Liu <sup>3</sup>,  
Andrew Tseng <sup>4</sup>, Wei Quan <sup>5</sup>, Neal Ryan <sup>6</sup>, David Brent <sup>6</sup>, Fuchiang R Tsui <sup>7</sup>

Affiliations + expand

PMID: 35007754 DOI: [10.1016/j.jbi.2021.103984](https://doi.org/10.1016/j.jbi.2021.103984)



# Impact of Social Determinants of Health



Only 20% of health impact factors coming from hospital care

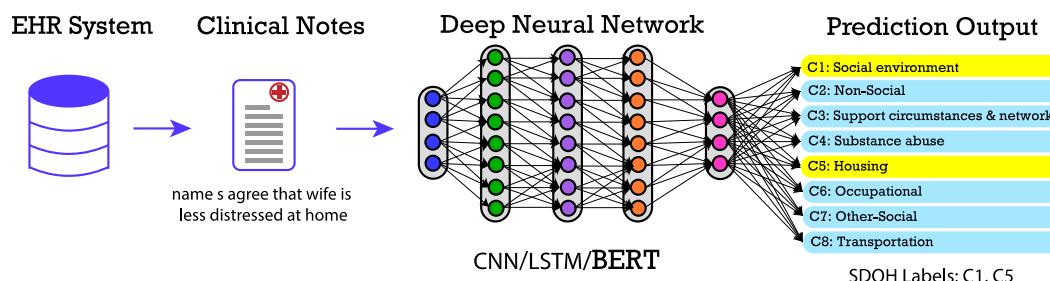


The remaining 80% of individual health impacting factors are from social determinants of health and health behavior

Table 3. Distribution of annotated social factor categories in the dataset. Note the percentages do not add up to 1, because each sentence can be assigned to one or more categories (428 sentences have more than one SDOH category).

Annotation Category	Count (% Frequency)	Category ID		CNN (95% CI)	LSTM (95% CI)	BERT (95% CI)
			Hamming Loss	0.097 (0.094, 0.100)	0.106 (0.102, 0.110)	0.095 (0.088, 0.102)
Social environment	1783 (50.9%)	C1	Macro F1	0.550 (0.527, 0.572)	0.555 (0.536, 0.574)	<b>0.642 (0.623, 0.662)*</b>
Non-Social	572 (16.3%)	C2	Micro F1	0.649 (0.639, 0.659)	0.632 (0.619, 0.644)	<b>0.690 (0.670, 0.710)*</b>
Support circumstances and networks	515 (14.7%)	C3	Macro Precision	<b>0.587 (0.567, 0.606)</b>	0.546 (0.533, 0.559)	0.583 (0.555, 0.612)
Substance abuse	385 (11.0%)	C4	Micro Precision	<b>0.667 (0.657, 0.676)</b>	0.620 (0.606, 0.634)	0.640 (0.613, 0.667)
Housing	236 (6.7%)	C5	Macro Recall	0.529 (0.502, 0.556)	0.574 (0.544, 0.603)	<b>0.756 (0.742, 0.770)*</b>
Occupational	191 (5.5%)	C6	Micro Recall	0.633 (0.619, 0.647)	0.644 (0.628, 0.659)	<b>0.750 (0.735, 0.765)*</b>
Other-Social	208 (5.9%)	C7	Macro AUC	0.854 (0.845, 0.864)	0.847 (0.834, 0.859)	<b>0.907 (0.899, 0.915)*</b>
Transportation	68 (1.9%)	C8	Micro AUC	<b>0.907 (0.903, 0.912)*</b>	0.865 (0.858, 0.872)	0.890 (0.879, 0.902)
			Macro Avg Precision	0.579 (0.557, 0.602)	0.556 (0.530, 0.581)	<b>0.698 (0.680, 0.715)*</b>
			Micro Avg Precision	0.695 (0.683, 0.708)	0.643 (0.624, 0.662)	<b>0.718 (0.694, 0.741)*</b>

\*: p<.05 according to Mann-Whitney U Tests.



# Evaluation and Implementation



Using one keyword, describe the top consideration in evaluation and implementation of AI models at point of care?

Explainability  
Accuracy  
Bias  
reproducibility

Usability  
Trials  
Quality

FAIRWorkflow  
transparency

MAYA  
Value  
safety

Impact

Privacy

Validity

Fairness

Equity

# Evaluation and Implementation

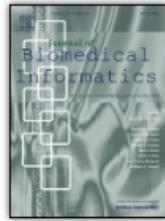


Review

› J Biomed Inform. 2022 Mar;127:103996. doi: 10.1016/j.jbi.2022.103996.

Epub 2022 Jan 15.

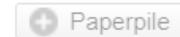
## Evaluating pointwise reliability of machine learning prediction



Giovanna Nicora <sup>1</sup>, Miguel Rios <sup>2</sup>, Ameen Abu-Hanna <sup>2</sup>, Riccardo Bellazzi <sup>3</sup>

Affiliations + expand

PMID: 35041981 DOI: 10.1016/j.jbi.2022.103996

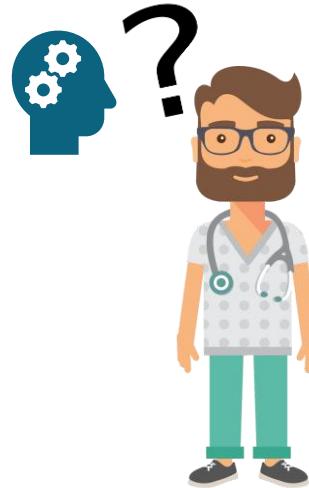


# Issues in the application of ML in Medicine



Poor generalization ability over time and/or across different population can cause a **lack of trust** in ML

In deployment, we need to implement methods to understand if we can trust the prediction on a new unlabeled case



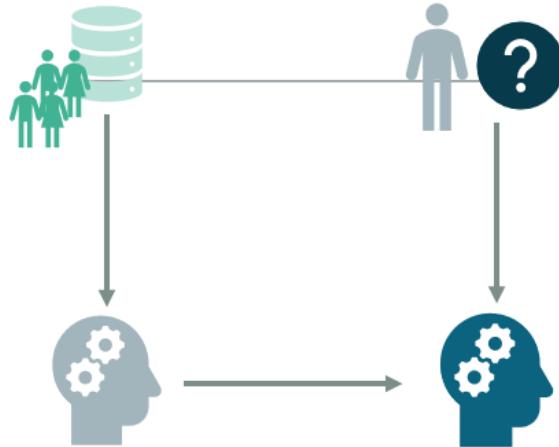
***Reliability: Degree of trust that the prediction is correct***

$$Rel(y_i): p(\hat{y}_i = y_i)$$

Images  
from [Pixabay](#)

Training data

New instance to be classified



Should we trust the prediction?



### Reliability Assessment

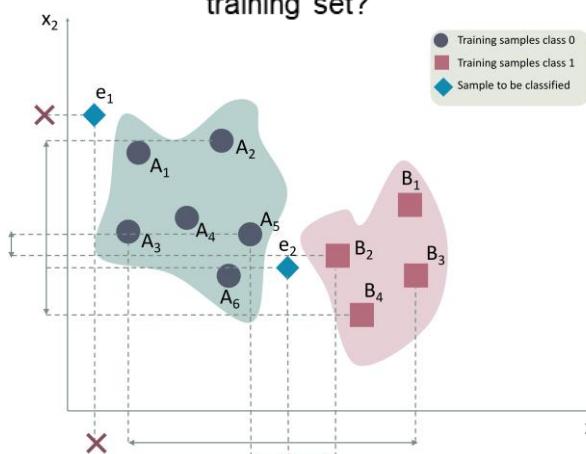


Local fit principle



Density principle

The new instance is close to the training set?

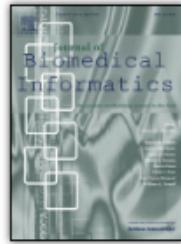


# Evaluation and Implementation



> J Biomed Inform. 2022 Aug;132:104107. doi: 10.1016/j.jbi.2022.104107. Epub 2022 Jun 7.

## Tell me something interesting: Clinical utility of machine learning prediction models in the ICU



Bar Eini-Porat <sup>1</sup>, Ofra Amir <sup>2</sup>, Danny Eytan <sup>3</sup>, Uri Shalit <sup>2</sup>

Affiliations + expand

PMID: 35688332 DOI: 10.1016/j.jbi.2022.104107

# Clinical Utility of Prediction Models in the ICU



ML models are transforming patient care into a collaborative **human-AI task**, yet prediction of patient-related events is mostly treated as a **standalone goal**.

**Tell Me Something Interesting:** a mixed methods study aimed at understanding clinicians' needs and expectations in the context of predictions at the ICU.

The findings identify several areas of focus where clinicians' needs deviate from current practice.

## Methods and Cohort

ICU practitioners from 4 different ICU wards (N=13)

Semi-structured exploratory interviews to identify clinical based ICU requirements

Quantitative tasks to evaluate clinical relevance of ML-based vital signs predictions.

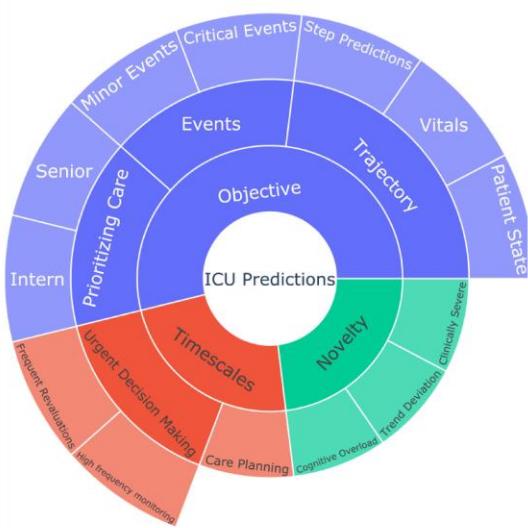


# Tell Me Something Interesting

## Exploratory Interviews

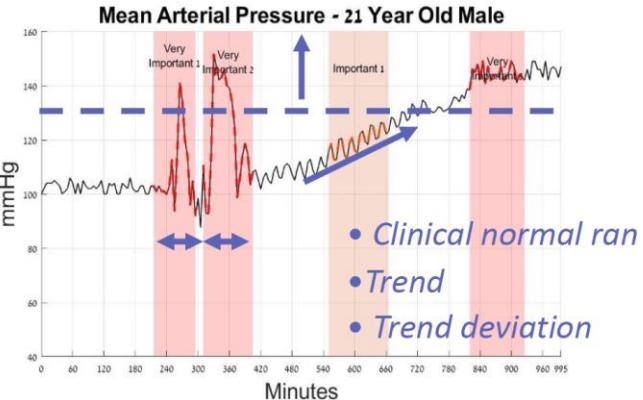
- ✓ Trajectory over critical events or exact prediction
- ✓ Two time scales – planning and urgent decision making
- ✓ NOVELTY!
- ✓ Underlying objective – prioritize care!

*Not all predictions are equal!*

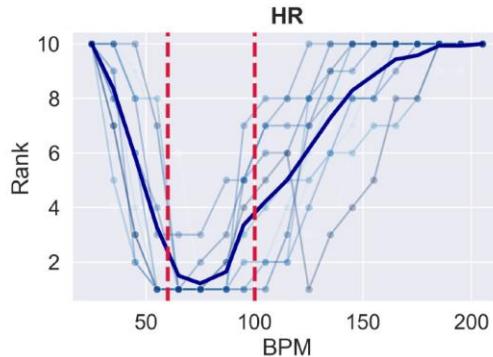


## Quantitative Tasks

Are we predicting what's important?



## Utility Curves



# Evaluation and Implementation



› J Am Med Inform Assoc. 2022 Sep 12;29(10):1810-1817. doi: 10.1093/jamia/ocac121.

## Tasks as needs: reframing the paradigm of clinical natural language processing research for real-world decision support



Asher Lederman <sup>1</sup>, Reeva Lederman <sup>1</sup>, Karin Verspoor <sup>2</sup>

Affiliations + expand

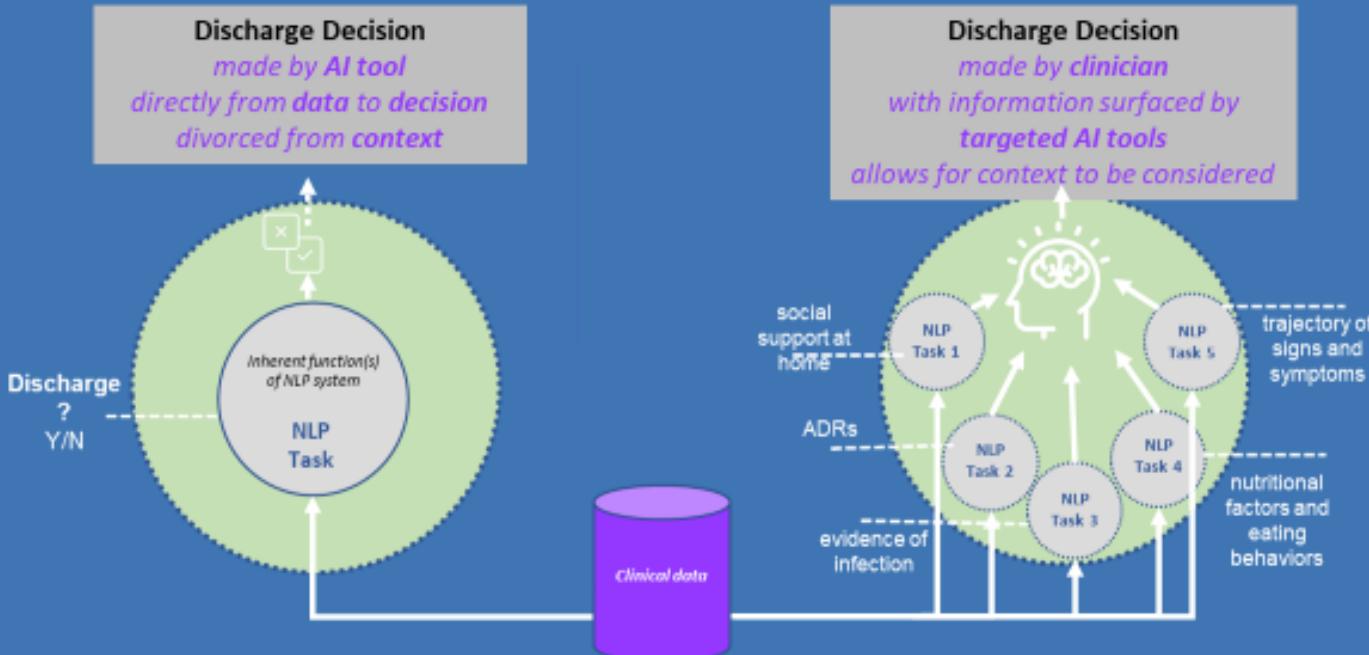
PMID: 35848784 PMCID: PMC9471702 DOI: 10.1093/jamia/ocac121





Friedman 2009. Friedman's Fundamental Theorem of Informatics  
 “A person working in partnership with an information resource is  
 ‘better’ than that same person unassisted”

## “Tasks as Decisions” vs “Tasks as Needs” Example: Hospital discharge decision



Shift focus of  
 AI/NLP tools from  
**end-to-end decision making**  
 to **supporting humans** to make  
 better clinical decisions

Human + AI >  
 Human >  
 AI

# Evaluation and Implementation



› J Am Med Inform Assoc. 2023 Feb 22;ocad006. doi: 10.1093/jamia/ocad006. Online ahead of print.

## Assessing the net benefit of machine learning models in the presence of resource constraints



Karandeep Singh <sup>1</sup> <sup>2</sup> <sup>3</sup> <sup>4</sup>, Nigam H Shah <sup>5</sup>, Andrew J Vickers <sup>6</sup>

Affiliations + expand

PMID: 36810659 DOI: [10.1093/jamia/ocad006](https://doi.org/10.1093/jamia/ocad006)

› J Biomed Inform. 2023 Mar;139:104319. doi: 10.1016/j.jbi.2023.104319. Epub 2023 Feb 13.

## APLUS: A Python library for usefulness simulations of machine learning models in healthcare



Michael Wornow <sup>1</sup>, Elsie Gyang Ross <sup>2</sup>, Alison Callahan <sup>3</sup>, Nigam H Shah <sup>4</sup>

Affiliations + expand

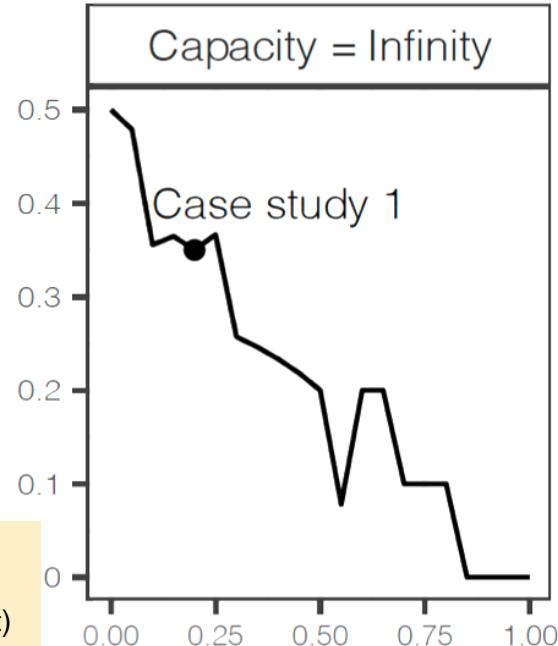
PMID: 36791900 DOI: [10.1016/j.jbi.2023.104319](https://doi.org/10.1016/j.jbi.2023.104319)

# Net benefit of ML under resource constraints

Patient #	Actually required ICU	Predicted risk (sorted from high to low)	Predicted to require ICU (Based on $P_t = 0.2$ )
1	Yes	0.8	Yes
2	Yes	0.7	Yes
3	No	0.6	Yes
4	Yes	0.5	Yes
5	Yes	0.3	Yes
6	No	0.2	Yes
7	No	0.1	No
8	No	0.1	No
9	Yes	0.05	No
10	No	0.01	No

## Case 1: no capacity constraints

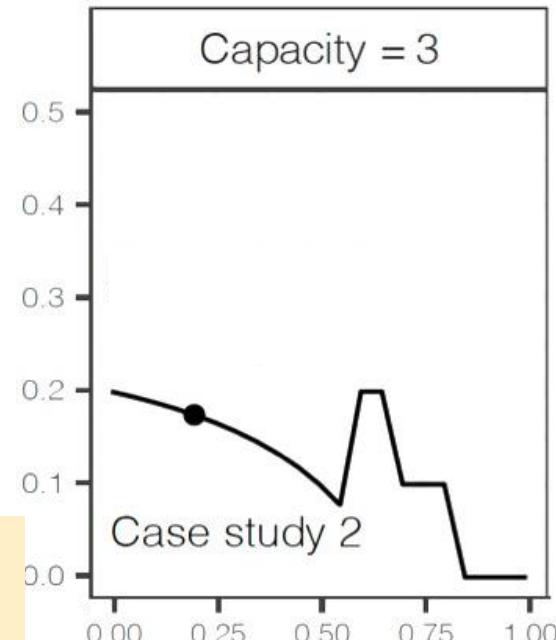
$$\begin{aligned} NB \times N &= TP - FP \times Pt / (1 - Pt) \\ &= 4 - 2 \times (0.2 / 0.8) \\ &= 4 - 0.5 = 3.5 \\ NB &= (NB \times N) / N = 3.5 / 10 = 0.35 \end{aligned}$$



# Net benefit of ML under resource constraints

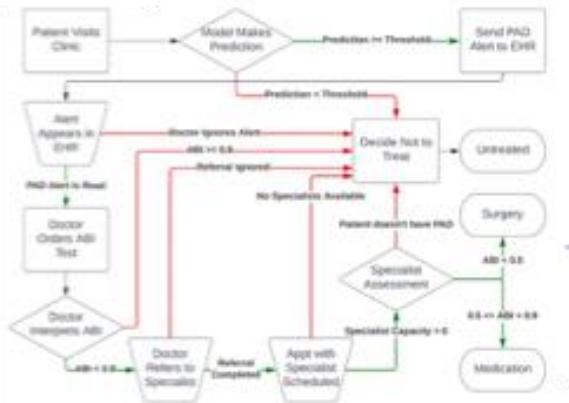
Patient #	Actually required ICU	Predicted risk (sorted from high to low)	Predicted to require ICU (Based on $P_t = 0.2$ )
1	Yes	0.8	Yes
2	Yes	0.7	Yes
3	No	0.6	Yes
4	Yes	0.5	Yes
5	Yes	0.3	Yes
6	No	0.2	Yes
7	No	0.1	No
8	No	0.1	No
9	Yes	0.05	No
10	No	0.01	No

**Case 2:  $P_t = 0.2$  and capacity = 3:**  
 $RNB \times N = TP_c - FP_c \times Pt/(1 - Pt)$   
 $= 2 - 1 \times (0.2 / 0.8)$   
 $= 2 - 0.25 = 1.75$   
 $RNB = 1.75/10 = 0.175$



## Model Outputs.csv

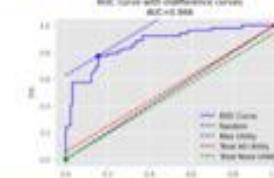
patient_id	prediction	label
0	0.01428571	0
1	0.01428571	0
2	0.21904762	0
3	0.09047619	0
4	0.0047619	0
5	0.01904762	0
6	0.0047619	0



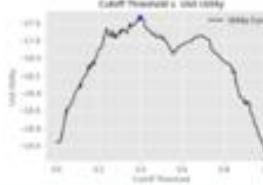
## APLUS Simulation Engine



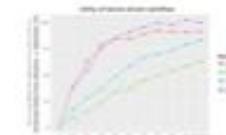
## Model Performance



## Theoretical Utility



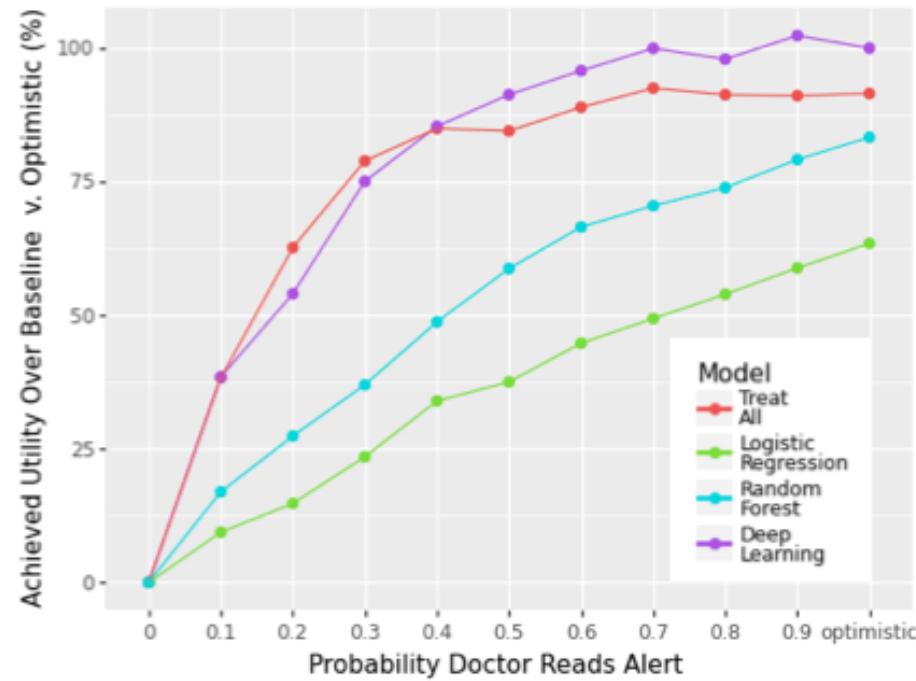
## Achieved Utility Under Workflow Constraints



# Python Library for Usefulness Simulations of ML Models

A **usefulness assessment** quantifies the **achievable utility** of an ML model under a **realistic** care delivery environment factoring in:

1. Model predictive performance
2. Workflow steps
3. Resource constraints
4. Utility of actions
5. Patient flow over time

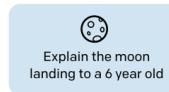


# Closing

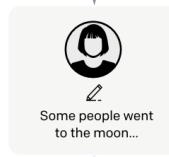
Step 1

**Collect demonstration data, and train a supervised policy.**

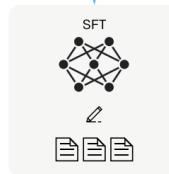
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



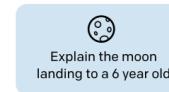
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

**Collect comparison data, and train a reward model.**

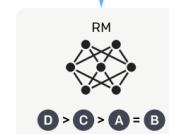
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



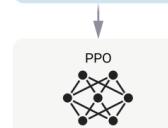
Step 3

**Optimize a policy against the reward model using reinforcement learning.**

A new prompt is sampled from the dataset.



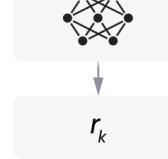
The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.





# When ChatGPT is asked by 3 of my team members

what topics should I include in an annual review of Artificial Intelligence and Data Science Year in Review?



An annual review of Artificial Intelligence and Data Science should cover the significant developments and trends in these fields in the past year. Here are some topics that you could include:

1. Advances in deep learning: Highlight the most significant developments in deep learning algorithms, such as new architectures, optimization techniques, and applications.
2. Natural Language Processing (NLP): Discuss the recent advances in NLP, including the development of new models, techniques for pretraining, and applications such as chatbots, voice assistants, and language translation.
3. Computer vision: Talk about the latest developments in computer vision, such as object detection, image segmentation, and video analysis.
4. Ethics and bias in AI: Address the ethical concerns surrounding AI and data science, including the use of biased data, algorithmic fairness, and privacy concerns.
5. Automation: Discuss the latest advancements in automation, including the use of robotic process automation (RPA), machine learning, and artificial intelligence.
6. Emerging technologies: Address new technologies that are emerging and their potential impact on AI and data science, such as blockchain, quantum computing, and edge computing.
7. Industry-specific applications: Discuss the latest trends and developments in AI and data science applications in specific industries, such as healthcare, finance, and manufacturing.
8. Data analysis: Talk about the latest techniques for data analysis, including machine learning algorithms, data visualization, and big data platforms.
9. Education and training: Address the need for education and training in AI and data science and discuss the latest developments in this area, including online courses and boot camps.
10. Future outlook: Discuss what the future holds for AI and data science and speculate on the latest trends and developments in these fields, such as the use of AI in robotics and autonomous vehicles, or the integration of AI with other emerging technologies.

Automation, RPA

Education and training

Emerging technologies

Hmmm ...

You are tech savvy ...



# When ChatGPT is asked by 3 of my team members

what topics should I include in an annual review of Artificial Intelligence and Data Science  
Year in Review

Here are some topics you may consider including in an annual review of Artificial Intelligence and Data Science:

1. Breakthroughs in AI Research: Highlight the latest advancements in AI research, such as new algorithms, models, and frameworks that have improved the accuracy and efficiency of AI systems.
2. Applications of AI and Data Science: Discuss the most significant applications of AI and data science in different fields such as healthcare, finance, transportation, and education. You can mention how AI is being used to develop new diagnostic tools, improve financial forecasting, or enhance transportation systems.
3. Ethics and Regulation: Address the ethical and regulatory issues that arose during the year regarding AI and data science, such as concerns about privacy, bias, and accountability. Mention any new laws or regulations that were passed to regulate the use of AI.
4. Industry Trends: Analyze the latest trends in the AI and data science industry, including new startups, mergers and acquisitions, and investments. Discuss any new tools or platforms that have been introduced to the market.
5. Natural Language Processing: Cover the progress made in the field of Natural Language Processing (NLP) during the year. Mention the latest models, algorithms, and techniques that have been developed to improve NLP, such as GPT-3, BERT, and RoBERTa.
6. Computer Vision: Address the latest advancements in computer vision, including new models for object detection, image classification, and segmentation. Mention any new datasets that have been released to the public for training computer vision models.
7. Machine Learning: Discuss the progress made in the field of machine learning during the year, including new models and algorithms that have improved the accuracy and efficiency of machine learning systems.
8. Robotics: Address the latest developments in the field of robotics, including advancements in machine learning and AI that are driving innovation in robotics. Mention any new robots or robotic systems that have been developed during the year.
9. Open Source Software: Discuss the latest advancements in open-source software for AI and data science, such as TensorFlow, PyTorch, and scikit-learn. Mention any new libraries or tools that have been developed to make it easier for developers to build and deploy AI models.
10. Future Outlook: Offer some thoughts on the future of AI and data science, including any trends or developments that you think will have a significant impact on the field in the coming years.

## Ethics and Regulation

Industry trends, startups, M&A, investments

Open source to deploy AI models

Hmmm ...

You are business savvy ...



# When ChatGPT is asked by 3 of my team members



what topics should I include in an annual review of Artificial Intelligence and Data Science Year in Review



Here are some topics that you may want to consider including in an annual review of Artificial Intelligence and Data Science Year in Review:



1. Major advancements in machine learning algorithms and deep learning models
2. Emerging trends in natural language processing (NLP) and speech recognition technology
3. Development of new tools and frameworks for data analysis and processing
4. Key breakthroughs in computer vision and image recognition
5. Applications of AI and data science in various industries such as healthcare, finance, and transportation
6. Ethical considerations surrounding the use of AI and data science
7. Research progress in areas such as reinforcement learning, causal inference, and explainable AI
8. Notable acquisitions, partnerships, and investments in the AI and data science space
9. Advances in autonomous systems and robotics
10. Impact of AI and data science on the job market and future of work.

Of course, this is not an exhaustive list, and you should tailor the review to your audience and goals.

RL, causal inference, explainable AI

Acquisitions + partnerships,  
investment

Impact on the job market

Hmmm ...

You are looking a job ...

# When ChatGPT is asked by 3 of my team members



Be careful when  
your supervisor  
asks you to ask  
ChatGPT!

# Can ChatGPT be used in healthcare?



No, it should not be used at all



Yes, it can only be used to help write administrative content such as emails to insurance companies or to patients



Yes, it can be used for any purpose

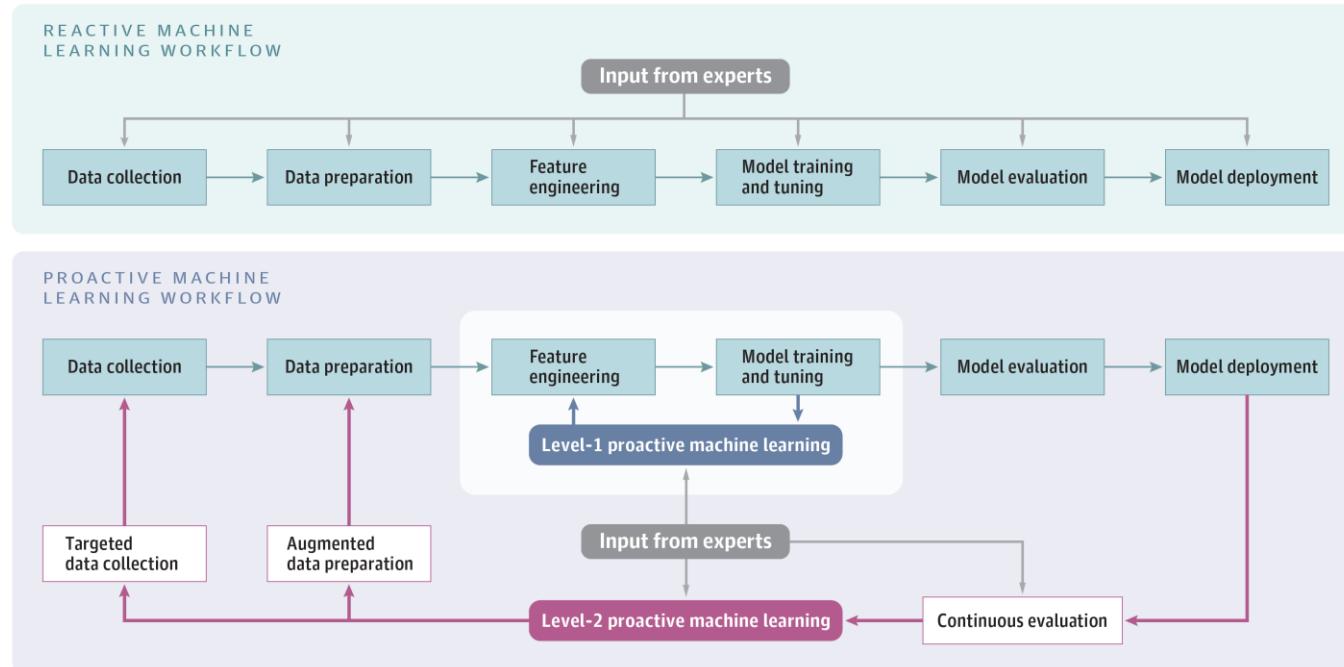


I don't know, it is too early to make a statement

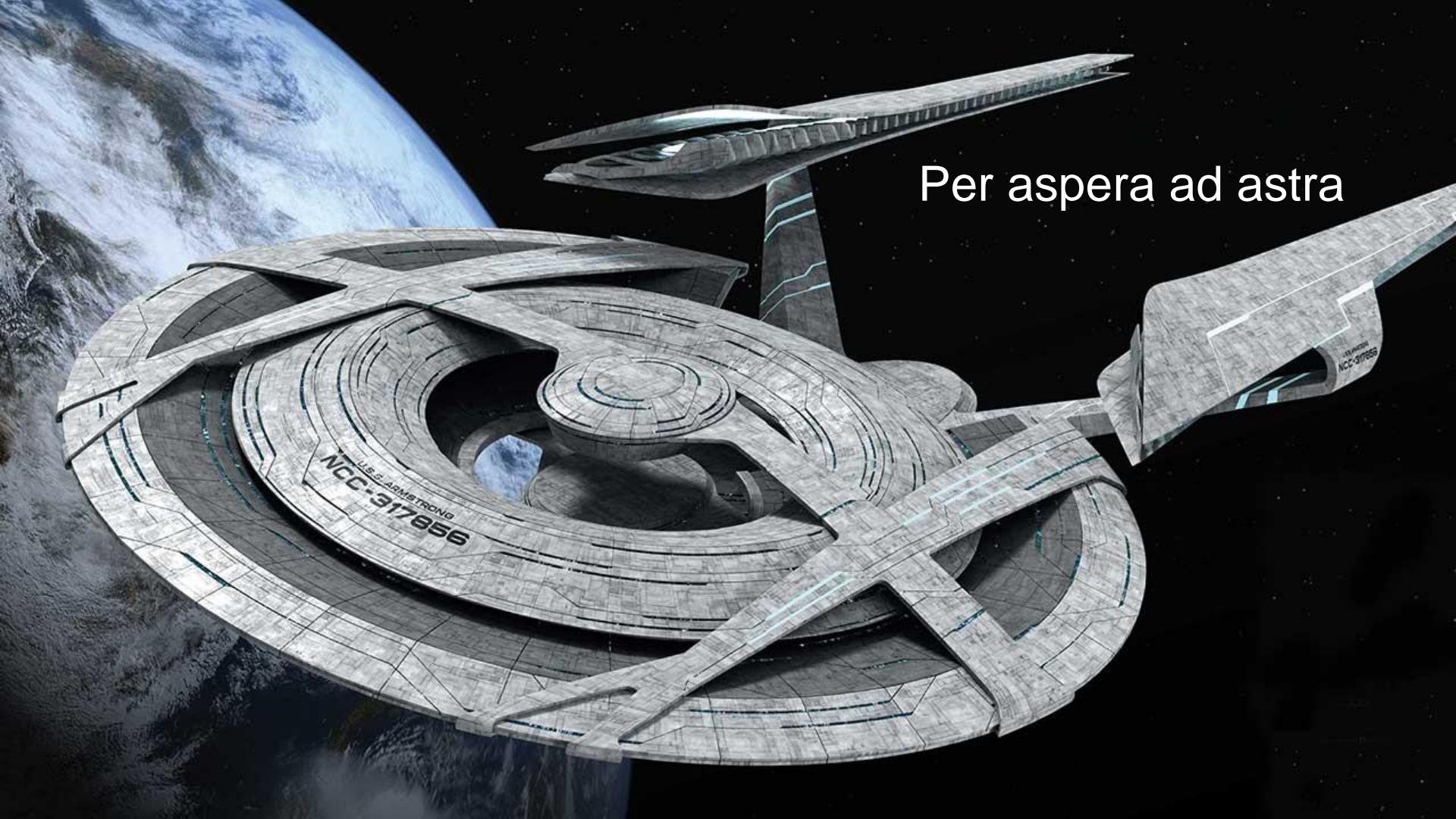


ⓘ Start presenting to display the poll results on this slide.

# Moving from reactive to proactive AI/ML and data science



Luo Y, Wunderink RG, Lloyd-Jones D. Proactive vs Reactive Machine Learning in Health Care: Lessons From the COVID-19 Pandemic. *JAMA*. 2022.



Per aspera ad astra