Weill Cornell Medicine
Institute of Artificial Intelligence for Digital Health

The WCM Institute of AI for Digital Health (AIDH)

Annual Report - 2024

Fei Wang. PhD. FACMI, FAMIA, FIAHSI.
Director, AIDH | Professor, Division of Health Informatics.
Department of Population Health Sciences.
Weill Cornell Medicine. Cornell University.
few2001@med.cornell.edu

Mission

Weill Cornell Medicine
Institute of Artificial Intelligence for Digital Health

Our mission is to advance **individualized health** and healthcare with **data-driven** insights derived from applying **artificial intelligence** technologies. The institute will establish a community of experts across multiple disciplines in **computational science** and **clinical medicine**, facilitate collaborations, promote **multi-disciplinary research**, and engage in **entrepreneurship** and technology **commercialization**.

Team

Team

Faculty



Research Scientists









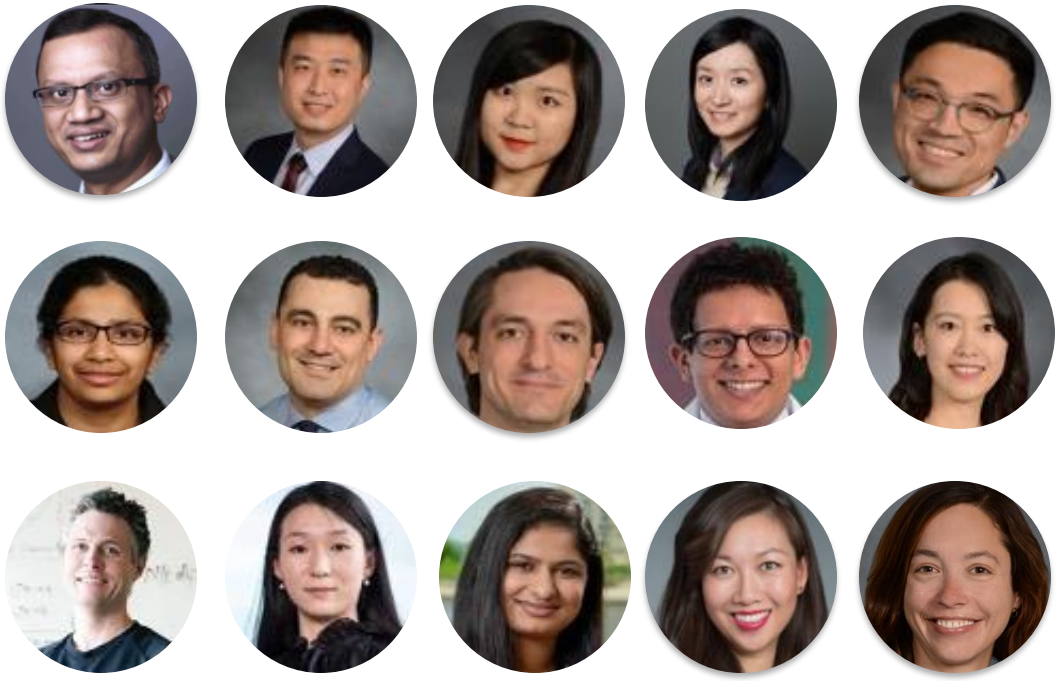
Graduate Students





Admin

Faculty Affiliates



Activities

Seminar Series

1x
month

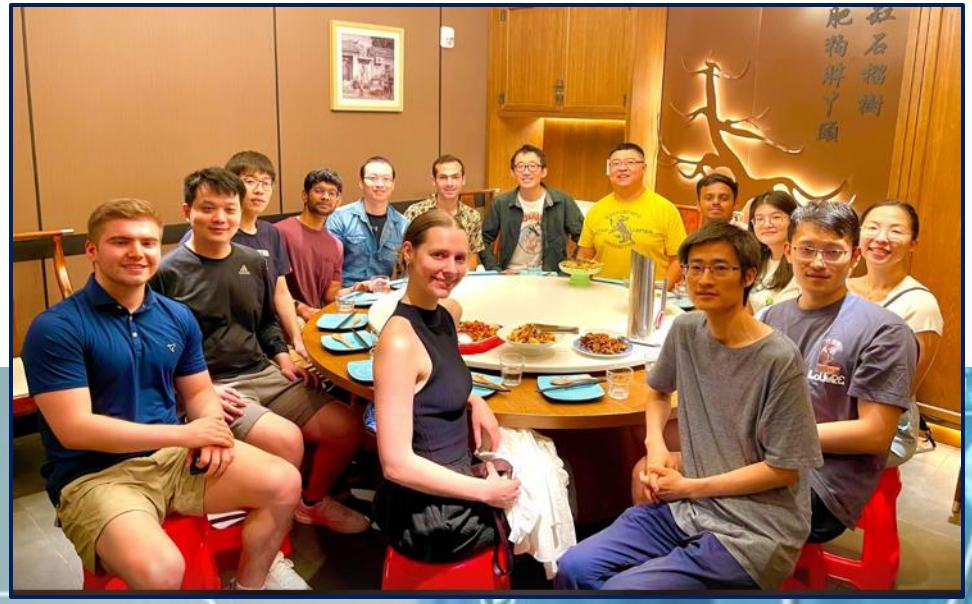
20
seminars

50+
participants per seminar

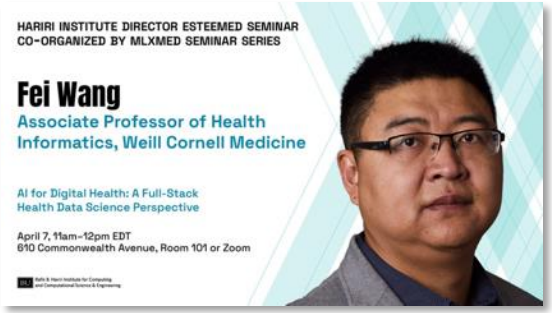
15
institutions represented

Gather Together

Weill Cornell Medicine
Institute of Artificial Intelligence for Digital Health



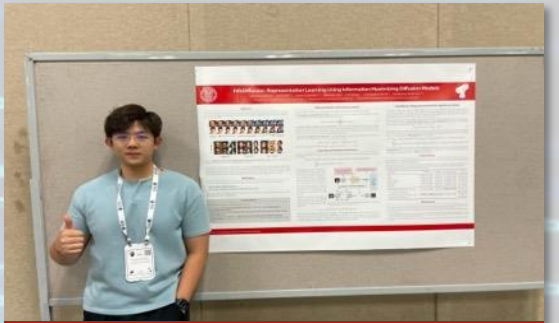
Invited Presentations



Conference Attendance

Chengxi Zhang, KDD

Weishen Pan KDD



Yingheng Wang, ICML

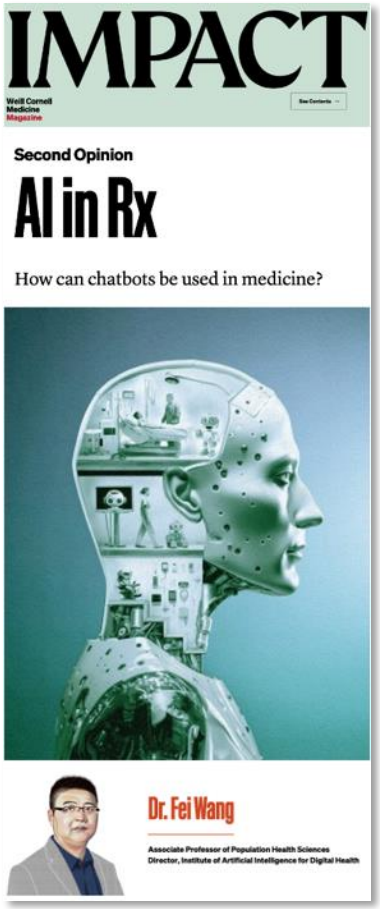
Daniel Adler, UbiComp/ISWC

Jacqueline Maasch, NeurIPS

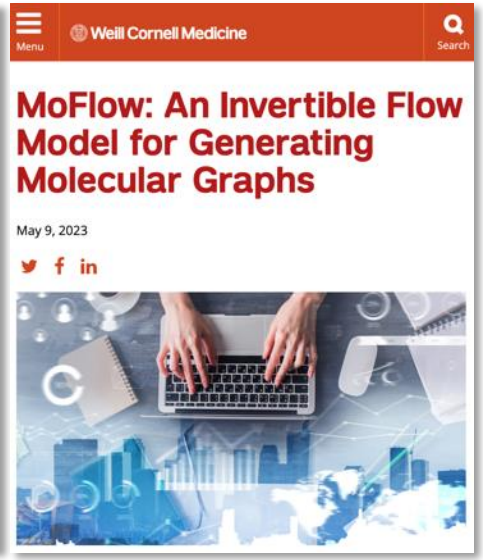
Zilong Bai ISGC

Press


WE'RE WORKING TO PREDICT THE FUTURE OF YOUR HEALTH.
At NewYork-Presbyterian, data scientists and doctors from Columbia and Weill Cornell Medicine are ushering in a new era of healthcare. They're combining their vast experience caring for diverse communities and pairing it with the latest in predictive analytics. Which means one day we can help you get ahead of a health issue before it becomes one—to stay amazing today and tomorrow.
STAY AMAZING
NewYork-Presbyterian
COLUMBIA
Weill Cornell Medicine
Learn more at nyp.org/stayamazing


IMPACT
Weill Cornell Medicine Magazine
Second Opinion
AI in Rx
How can chatbots be used in medicine?
Dr. Fei Wang
Associate Professor of Population Health Sciences
Director, Institute of Artificial Intelligence for Digital Health


Menu
Weill Cornell Medicine
Search
Dr. Fei Wang Provides Insight on the Use of Artificial Intelligence in Healthcare Diagnostics
August 22, 2023


Menu
Weill Cornell Medicine
Search
MoFlow: An Invertible Flow Model for Generating Molecular Graphs
May 9, 2023


360Dx
Menu
With Pathology at Vanguard, AI Infiltrates Healthcare Diagnostics Space
Jul 25, 2023 | Greg Cima


The Atlantic
Sign In
HEALTH
Doctors Still Struggle to Diagnose a Condition That Kills More Americans Than Stroke
Can computers crack the code of sepsis?


HealthDay
Home
Health A to Z
News For Medical Professionals
Video
Infectious Disease
Blood Tests for Long COVID Could Lead to Better Treatments

Research

Publications

52
Papers

5100+
Citations

2100+
Tweets

190+
News Outlets

nature medicine

nature COMMUNICATIONS

JAMA Pediatrics

npj Digital Medicine

Annals of Internal Medicine

Cell Reports Methods

nature medicine

Article

https://doi.org/10.1038/s41591-022-02116-3

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

Received: 8 June 2022

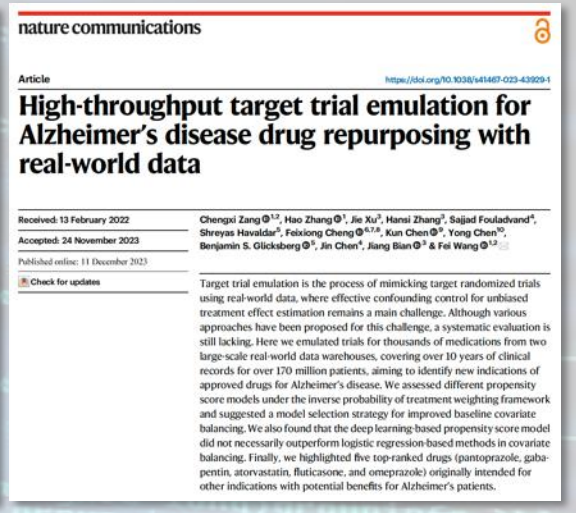
Accepted: 2 November 2022

Published online: 1 December 2022

Check for updates

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

The post-acute sequelae of SARS-CoV-2 infection (PASC) refers to a broad spectrum of symptoms and signs that are persistent, exacerbated or newly incident in the period after acute SARS-CoV-2 infection. Most studies have examined these conditions individually without providing evidence on co-occurring conditions. In this study, we leveraged the electronic health record data of two large cohorts, INSIGHT and OneFlorida+, from the national Patient-Centered Clinical Research Network. We created a development cohort from INSIGHT and a validation cohort from OneFlorida+ including 20,881 and 13,724 patients, respectively, who were SARS-CoV-2 infected, and we investigated their newly incident diagnoses 30–180 days after a documented SARS-CoV-2 infection. Through machine learning analysis of over 137 symptoms and conditions, we identified four reproducible PASC subphenotypes, dominated by cardiac and renal (including 33.75% and 25.43% of the patients in the development and validation cohorts); respiratory, sleep and anxiety (32.75% and 38.48%); musculoskeletal and nervous system (23.37% and 23.35%); and digestive and respiratory system (10.14% and 12.74%) sequelae. These subphenotypes were associated with distinct patient demographics, underlying conditions before SARS-CoV-2 infection and acute infection phase severity. Our study provides insights into the heterogeneity of PASC and may inform stratified decision-making in the management of PASC conditions.



nature communications

Article

https://doi.org/10.1038/s41467-023-43929-1

High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data

Received: 13 February 2022

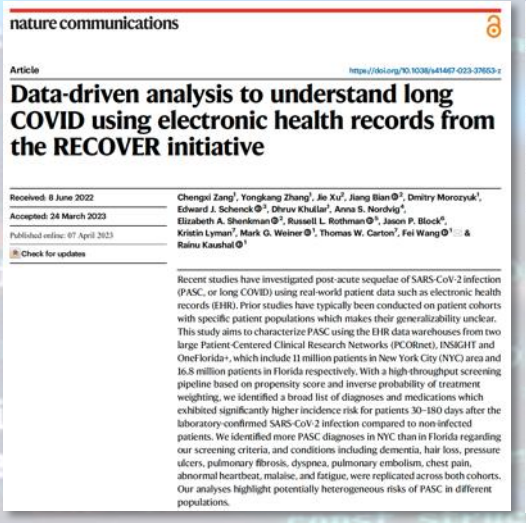
Accepted: 24 November 2023

Published online: 11 December 2023

Check for updates

Chengxi Zang[1,2], Hao Zhang[1], Jie Xu[3], Hansi Zhang[3], Sajjad Fouladvand[4], Shreyas Havaldar[5], Feixiong Cheng[6,7,8], Kun Chen[9], Yong Chen[10], Benjamin S. Glicksberg[5], Jin Chen[4], Jiang Bian[2] & Fei Wang[1,2] ✉

Target trial emulation is the process of mimicking target randomized trials using real-world data, where effective confounding control for unbiased treatment effect estimation remains a main challenge. Although various approaches have been proposed for this challenge, a systematic evaluation is still lacking. Here we emulated trials for thousands of medications from two large-scale real-world data warehouses, covering over 10 years of clinical records for over 170 million patients, aiming to identify new indications of approved drugs for Alzheimer's disease. We assessed different propensity score models under the inverse probability of treatment weighting framework and suggested a model selection strategy for improved baseline covariate balancing. We also found that the deep learning-based propensity score model did not necessarily outperform logistic regression-based methods in covariate balancing. Finally, we highlighted five top-ranked drugs (pantoprazole, gabapentin, atorvastatin, fluticasone, and omeprazole) originally intended for other indications with potential benefits for Alzheimer's patients.



nature communications

Article

https://doi.org/10.1038/s41467-023-37853-z

Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative

Received: 8 June 2022

Accepted: 24 March 2023

Published online: 07 April 2023

Check for updates

Chengxi Zang[1], Yongkang Zhang[1], Jie Xu[2], Jiang Bian[2], Dmitry Morozyuk[3], Edward J. Schenck[3], Dhruv Khullar[3], Anna S. Nordvig[4], Elizabeth A. Shenkman[2], Russell L. Rothman[5], Jason P. Block[6], Kristin Lyman[7], Mark G. Weiner[1], Thomas W. Carton[7], Fei Wang[1] ✉ & Rainu Kaushal[1]

Recent studies have investigated post-acute sequelae of SARS-CoV-2 infection (PASC, or long COVID) using real-world patient data such as electronic health records (EHR). Prior studies have typically been conducted on patient cohorts with specific patient populations which makes their generalizability unclear. This study aims to characterize PASC using the EHR data warehouses from two large Patient-Centered Clinical Research Networks (PCORnet), INSIGHT and OneFlorida+, which include 11 million patients in New York City (NYC) area and 16.8 million patients in Florida respectively. With a high-throughput screening pipeline based on propensity score and inverse probability of treatment weighting, we identified a broad list of diagnoses and medications which exhibited significantly higher incidence risk for patients 30–180 days after the laboratory-confirmed SARS-CoV-2 infection compared to non-infected patients. We identified more PASC diagnoses in NYC than in Florida regarding our screening criteria, and conditions including dementia, hair loss, pressure ulcers, pulmonary fibrosis, dyspnea, pulmonary embolism, chest pain, abnormal heartbeat, malaise, and fatigue, were replicated across both cohorts. Our analyses highlight potentially heterogeneous risks of PASC in different populations.

Research

JAMA Pediatrics | Original Investigation

Patterns of Social Determinants of Health and Child Mental Health, Cognition, and Physical Health

Yunyu Xiao, PhD; J. John Mann, MD; Julian Chun-Chung Chow, PhD; Timothy T. Brown, PhD; Lonnie R. Snowden, PhD; Paul Siu-Fai Yip, PhD; Alexander C. Tsai, MD, PhD; Yu Hou, PhD; Jyotishman Pathak, PhD; Fei Wang, PhD; Chang Su, PhD

IMPORTANCE Social determinants of health (SDOH) influence child health. However, most previous studies have used individual, small-set, or cherry-picked SDOH variables without examining unbiased computed SDOH patterns from high-dimensional SDOH factors to investigate associations with child mental health, cognition, and physical health.

OBJECTIVE To identify SDOH patterns and estimate their associations with children's mental, cognitive, and physical developmental outcomes.

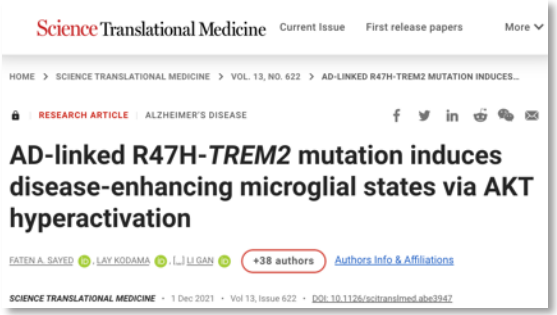
DESIGN, SETTING, AND PARTICIPANTS This population-based cohort study included children aged 9 to 10 years at baseline and their caregivers enrolled in the Adolescent Brain Cognitive Development (ABCD) Study between 2016 and 2021. The ABCD Study includes 21 sites across 17 states.

EXPOSURES Eighty-four neighborhood-level, geocoded variables spanning 7 domains of SDOH, including bias, education, physical and health infrastructure, natural environment, socioeconomic status, social context, and crime and drugs, were studied. Hierarchical agglomerative clustering was used to identify SDOH patterns.

Funding

Involved Projects Totaling 57M, AIDH Portion Total: 9.75M

Project Highlights


Science Translational Medicine Current Issue First release papers More

HOME > SCIENCE TRANSLATIONAL MEDICINE > VOL. 13, NO. 622 > AD-LINKED R47H-TREM2 MUTATION INDUCES...

RESEARCH ARTICLE | ALZHEIMER'S DISEASE

AD-linked R47H-*TREM2* mutation induces disease-enhancing microglial states via AKT hyperactivation

FATEN A. SAYED, LAY KODAMA, [...] LI GAN +38 authors Authors Info & Affiliations

SCIENCE TRANSLATIONAL MEDICINE • 1 Dec 2021 • Vol 13, Issue 622 • DOI: 10.1126/scitranslmed.abe3947

R01AG076448
PI: Feixiong Cheng, **Fei Wang**, Li Gan
2022-2027


AJPH

A PUBLICATION OF THE AMERICAN PUBLIC HEALTH ASSOCIATION

Sign In

Search

Home Articles Authors Subscri

Home » AJPH » May 2022

Prevalence of Alzheimer's and Related Dementia Diseases and Risk Factors Among Transgender Adults, Florida, 2012-2020

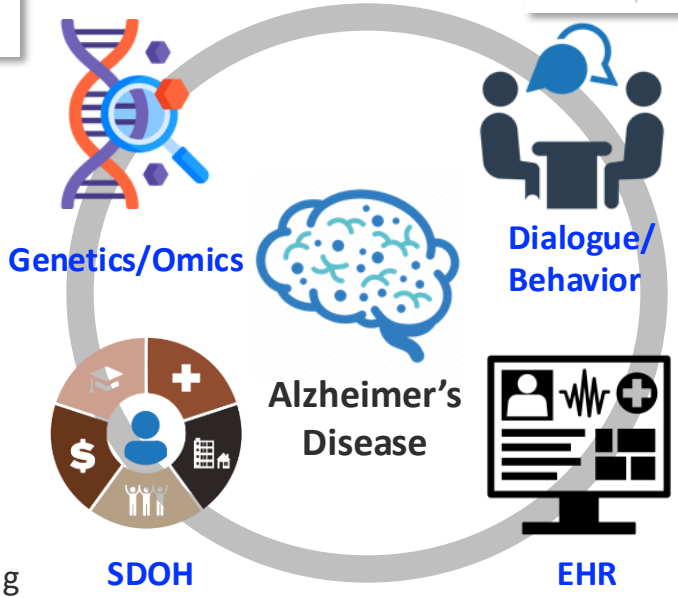
Yi GuoPhD, Qian LiMS, Xi YangPhD, Michael S. JaffeeMD, Yonghui WuPhD, Fei WangPhD, and Jiang BianPhD

[+] Author affiliations, information, and correspondence details

Accepted: January 05, 2022 Published Online: April 13, 2022

R01AG080624
PI: Jiang Bian, Yi Guo, **Fei Wang**
2023-2028

RF1AG084178
PI: Hui Hu, Jiang Bian, **Fei Wang**
2023-2028




scientific reports

Explore content About the journal Publish with us

nature > scientific reports > articles > article

Article | Open Access | Published: 31 March 2020

Scalable diagnostic screening of mild cognitive impairment using AI dialogue agent

Fengyi Tang, Ikechukwu Uchendu, Fei Wang, Hiroko H. Dodge & Jiayu Zhou

Scientific Reports 10, Article number: 5732 (2020) | Cite this article

3362 Accesses | 6 Citations | 30 Altmetric | Metrics

RF1AG072449
PI: Jiayu Zhou, **Fei Wang**, Hitoko Dodge
2021-2024


nature communications

Explore content About the journal Publish with us

nature > nature communications > articles > article

Article | Open access | Published: 11 December 2023

High-throughput target trial emulation for Alzheimer's disease drug repurposing with real-world data

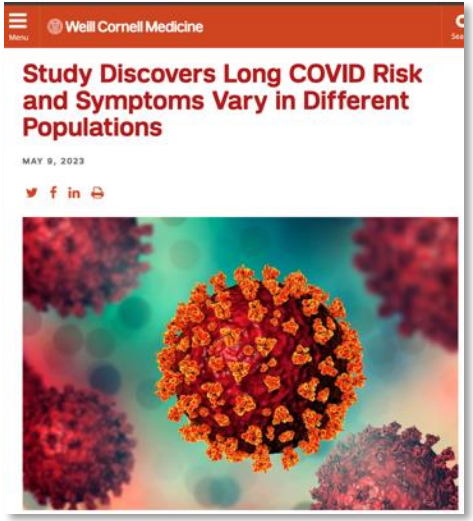
Chengxi Zang, Hao Zhang, Jie Xu, Hansi Zhang, Sajjad Fouladvand, Shreyas Havaldar, Feixiong Cheng, Kun Chen, Yong Chen, Benjamin S. Glicksberg, Jin Chen, Jiang Bian & Fei Wang

Nature Communications 14, Article number: 8180 (2023) | Cite this article

R01AG076234
PI: **Fei Wang**, Bian Jiang
2022-2027

R01AG080991
PI: **Fei Wang**, Bian Jiang
2023-2028

Large Initiative Participation: RECOVER PCORnet Adult EHR Study



PI: Dr. Rainu Kaushal

Our role: Lead the efforts on machine learning for deriving insights of long COVID from large scale EHRs

https://recovercovid.org/

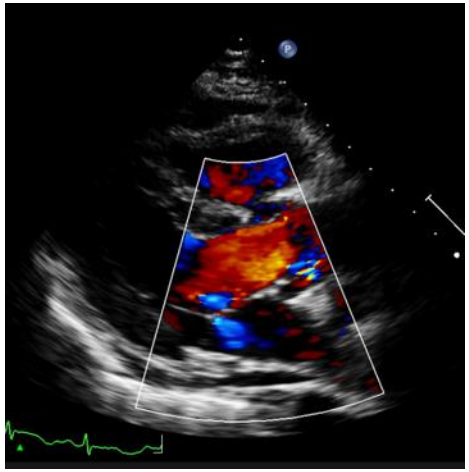
Large Initiative Participation: The NYP Cardiology AI Initiative

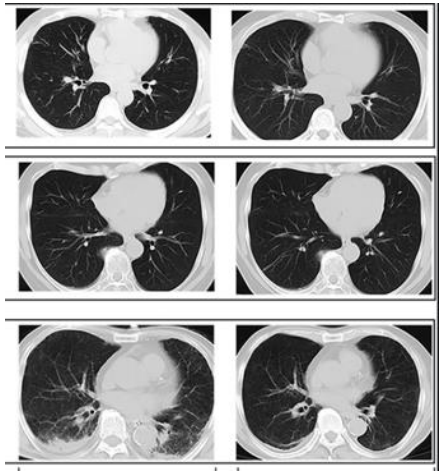
$15M Collaboration Aims to Advance Cardiovascular Care with Analytics, AI

NewYork-Presbyterian will contribute $15 million in funding to support a new initiative aimed at predicting and preventing heart failure and improving cardiovascular health interventions overall.

https://healthitanalytics.com/news/15m-collaboration-aims-to-advance-cardiovascular-care-with-analytics-ai

PI: Dr. Deborah Estrin & Kavita Bala
NYP Contact: Ashley Beecy
Our role: Co-lead the efforts of machine learning for predicting heart function from echocardiograms and chest CTs





Education/Training

AI in Medicine II

This class will teach students more advanced topics on AI in medicine. It requires students to have taken the AI in medicine I class. The contents of the class cover generalizability of AI models, computational fairness, model interpretation and explanation, privacy and security, federated learning, multi-modal learning, generative AI, causal inference, target trial emulation. The students will be asked to do a final project with teams based on the contents taught in the class, and python programming will be needed for doing the project.

Machine Learning for Health

This course introduces students the various types of health data, including patient clinical records, medical images, physiological and vital signals from wearable sensors, multi-omics, etc. and how to use machine learning algorithms to analyze these data and help with real world health problems such as patient screening, risk modeling, disease subtyping and precision medicine. The technical topics to be introduced in this class include classic supervised and unsupervised learning, network analysis, probabilistic modeling, deep learning, transfer learning, algorithmic fairness and interpretability. We will also invite clinicians or researchers working in the health industry to deliver guest lecturers in the class. The students will gain hands-on experience on analyzing real world health data during course assignments and projects.

AI for Healthcare System Engineering

The purpose of this class is to teach students the fundamentals of artificial intelligence (AI) technologies and how they can be applied in various healthcare system engineering problems. We will start introducing students the basic concepts of artificial intelligence (AI) and its history, as well as the various types of healthcare data and their representations. We will then introduce conventional AI technologies including supervised learning for tasks like clinical risk prediction and computer assisted diagnosis, unsupervised learning methods for subtype identification and pattern discovery. Next we will introduce deep learning methods, including the basic perceptron and feedforward neural networks for standard vectorized data, convolutional neural networks for analyzing medical images, recurrent neural networks and transformer for analyzing event sequences and temporal signals, and graph neural networks for analyzing networks and relational data. The class includes both lectures introducing algorithms and theories, and programming exercises to get hands-on experience on implementing these algorithms with Python. After the class the students will be able to understand the basic concept of AI and solve the real world healthcare system engineering problems with the learned knowledge.

Alumni & Faculty Mentorship

PhD Students

Postdoc Fellows

Junior Faculty K-Awardees

Tech Transfer

TrialGenie: A Multi-Agent System for Intelligent Clinical Trial Design

Fei Wang, Ph.D.
Professor of Population Health Sciences, Weill Cornell Medical College

Patents

Patents: Provisional Application Filed

Technology Overview

- **The Technology:** A multiagent system which allows insights to be extracted from real world data to perform target trial emulation
- Distinct computational nodes—a “Supervisor,” “Trialist,” “Clinician,” “Informatician,” and “Statistician”—handle specialized tasks such as extracting clinical trial information or mapping trial information to electronic health records
- The platform’s modular, agent-based design ensures consistent interplay between clinical domain expertise and analytics rigor
- **PoC Data:** The platform was used to emulate a historical trial evaluating the effectiveness of corticosteroids in managing sepsis within the ICU, identifying an average treatment effect and hazard ratio consistent with effects from the actual randomized controlled trial
- **Applications:** Post-market studies assessing real-world effectiveness, identification of label expansion opportunities, clinical trial design and planning, adaptive clinical trial design, strategic decision making and portfolio prioritization

Supporting Data / Figures

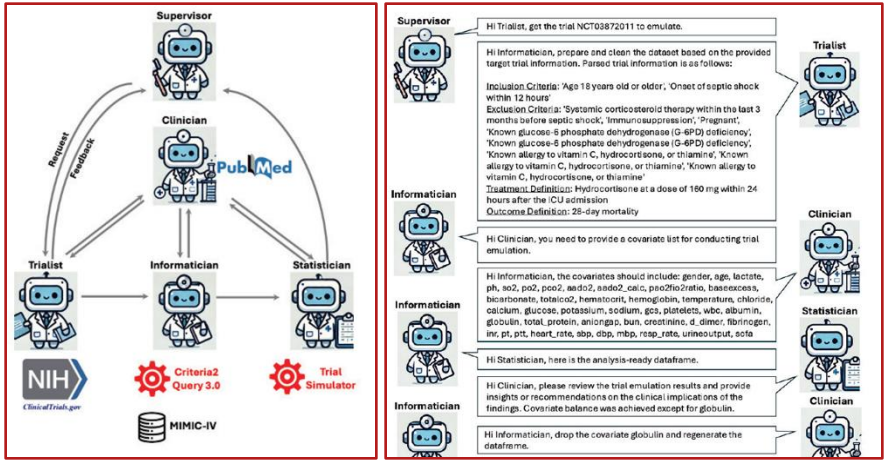


Figure 1: Left: Overview of agent interactions within the emulation framework **Right:** Flowchart illustrating the communication process and task delegation among the agents in the multi-agent system for clinical trial emulation.

Non-Invasive Device for Voice Restoration After Laryngectomy

Anaïs Rameau, M.D.
Assistant Professor of Otolaryngology, Weill Cornell Medical College

Technology Overview

- **The Technology:** A novel, personalized device for voice restoration using machine learning applied to surface EMG (sEMG) signal
- The inventors have created a tailored device to conform to a patient's unique anatomy with sensors on the articulatory muscles of the face and neck
- The device detects the sEMG signals and applies a predictive machine-learning model to translate silent speech into words
- **PoC Data**: The inventors collected data using this device from a laryngectomy patient silently articulating 'Tedd' and 'Ed'
- The team trained a predictive model for automatic speech recognition of these words, which had an 86.4% word recognition accuracy
- **Applications:**
  - Voice restoration for laryngectomy patients
  - Silent speech recognition for noisy or difficult environments

Patents & Publications

Patents: US Application Filed
Publication: Rameau. *Head N*eck. 2020.

Supporting Data / Figures

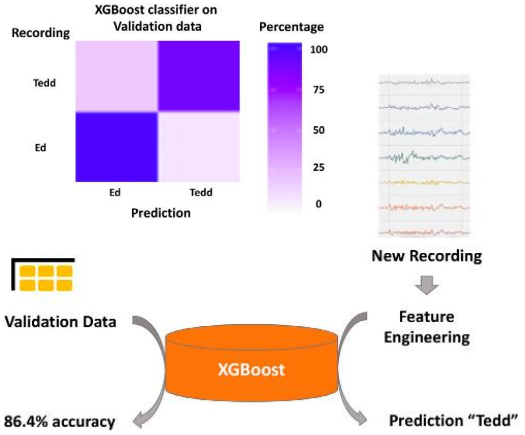
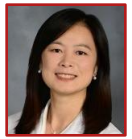


Figure 1: Left: Render of the device in position on patient **Right:** Schematic and results of machine learning model 'XGBoost' applied to validation data.

Footnotes

Integrated AI Platform for Early Preeclampsia Detection and Risk Assessment



Zhen Zhao, Ph.D.
Professor of Clinical Pathology and Laboratory Medicine,
Weill Cornell Medical College

Technology Overview

- **The Technology:** An AI-driven system for early prediction of preeclampsia risk that integrates routine clinical data with machine learning algorithms.
- The platform integrates diverse clinical inputs (including standard laboratory tests, maternal characteristics, and optional specialized biomarkers) to enable risk assessment across all healthcare settings, from hospitals to resource-limited environments.
- The system incorporates point-of-care testing capabilities and provides automated, user-friendly results without requiring specialized expertise
- **Applications:** The platform enables routine screening for early risk assessment and intervention, while supporting remote patient monitoring through telehealth integration
- Its flexibility allows deployment across various healthcare settings, from sophisticated hospital systems to resource-limited environments, making standardized preeclampsia risk assessment more widely accessible

Patents & Publications

Patents: Provisional Filed

Supporting Data / Figures

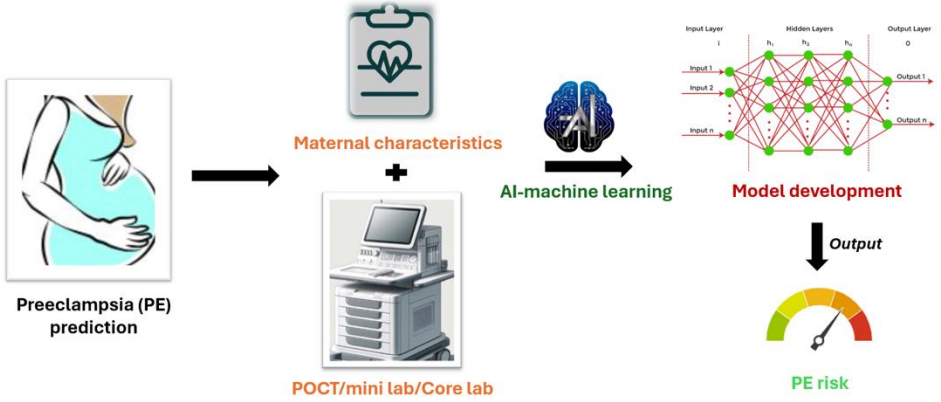


Figure 1: Overview of AI-driven preeclampsia prediction system workflow. The platform combines maternal characteristics with laboratory data from various testing options (POCT/mini lab/core lab) to generate a preeclampsia risk assessment through machine learning algorithms.

NerveAI: An ML-enabled Platform for Identification of Nerve Pain

Lisa Gfrerer, M.D.
Assistant Professor of Surgery (Plastic Surgery), Surgery, Weill Cornell Medical College

Technology Overview

- **The Technology:** Platform for screening patients for nerve pain using a digital 3D model of the head on which patients draw their pain
- The platform leverages AI-based pattern recognition to automatically evaluate pain drawings to diagnose nerve pain and identify patients that are candidates for headache surgery
- A prototype of the platform has been developed and trained on 1,300 3D pain drawings
- **PoC Data:** The highest performing model, a multilayer perceptron (MLP) model, distinguished nerve pain from other types of head and neck pain with an AUROC of 0.879, precision of 0.943, specificity of 0.611, and sensitivity of 0.640
- Another model, XGBoost, performed exceptionally well in detecting different types of nerve pain such Trigeminal Neuralgia (AUROC: 0.954), occipital nerve pain (AUROC: 0.928), and frontal nerve pain (AUROC: 0.930)

Patents & Publications

Patents: Provisional Filed

Publications: Chartier et al. *Plast Reconstr Surg.* 2023. Gfrerer et al. *Plast Reconstr Surg.* 2020.

Supporting Data / Figures

http://aidh.weill.cornell.edu/

few2001@med.cornell.edu