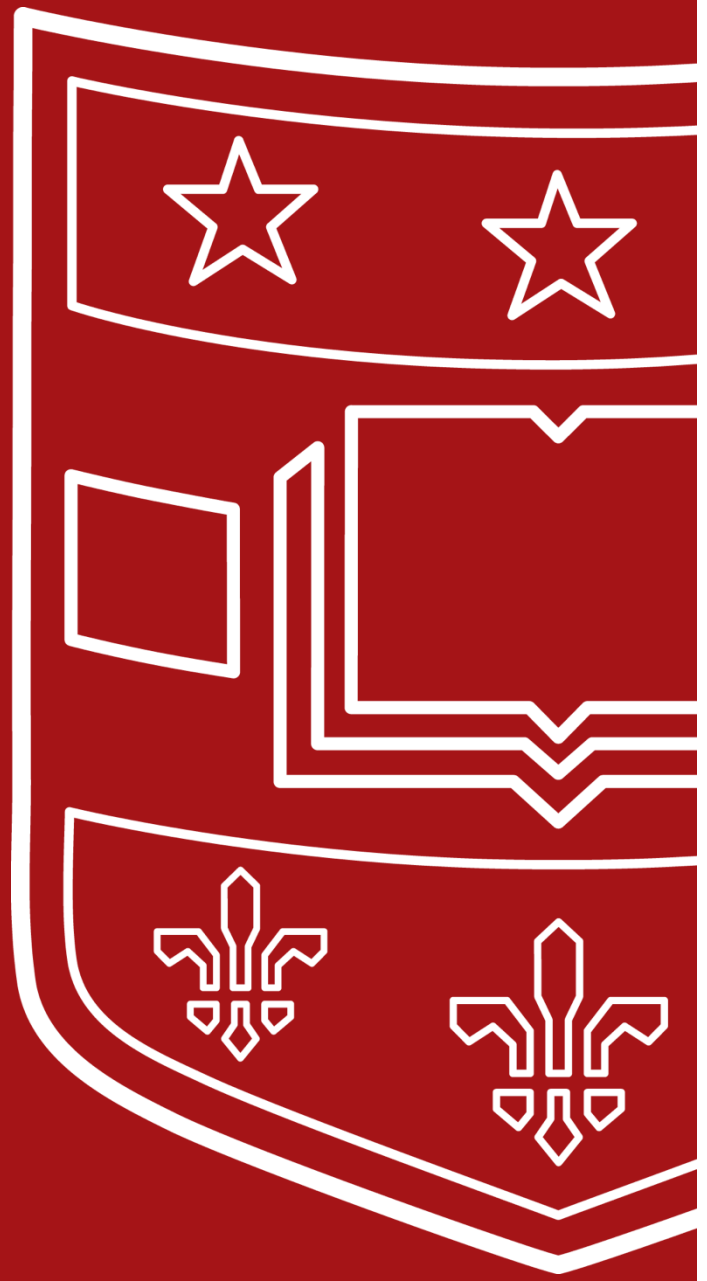
AI Across Medicine

Opportunities for Radiation Oncology

Chenyang Lu

Director, AI for Health Institute (AIHealth)

Fullgraf Professor of Computer Science, Anesthesiology, Medicine, Neurosurgery



Washington University in St. Louis

AI for Health

- Powerful data-driven tools for health
  - **Phenotype** complex diseases and population health
  - **Predict** health outcomes, risk factors, and treatment responses
  - **Guide** clinical decisions and public health policies
  - **Streamline** healthcare processes

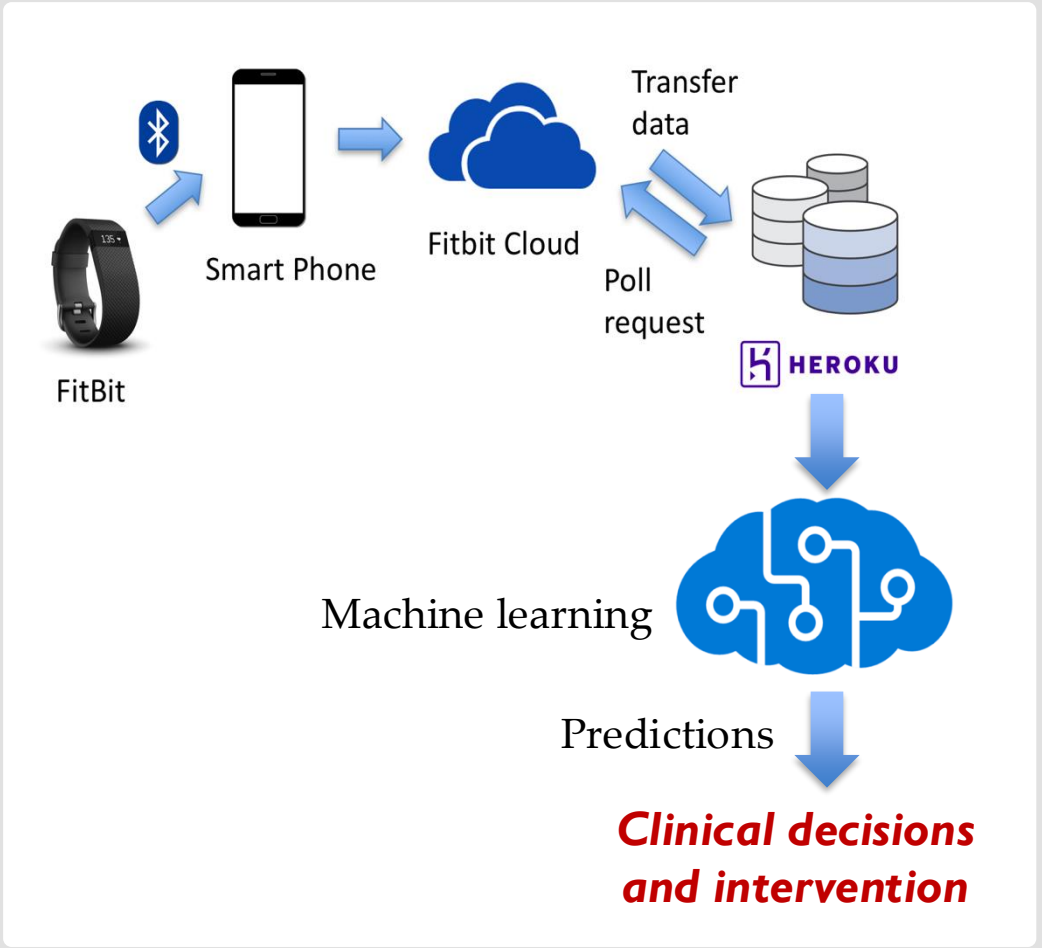
- Extract knowledge from complex health data
  - **Health care**: electronic health record, genomics, clinical notes, imaging
  - **Public health**: public health record, surveys, social determinants of health
  - **Mobile health**: smart phones, wearables

AI + Health

- Health
  - Health expertise and domain knowledge
  - **Struggle with multi-modal, high dimensional, noisy data**
- AI
  - New techniques are developed at an explosive rate
  - **Scarcely applied to health care and public health**
- AI + Health
  - **Solve significant health problems with advanced AI**

Decision Support for Pancreatic Surgery

- Pancreatic cancer has a 5-year survival rate less than 5%.
- Surgery is the only cure but commonly followed by complications.
  - Pancreatectomy: 90-day mortality rate over 4% and serious morbidity rates over 20%.
- Predict postoperative complications before surgery
  - **Decision support**: suitability for surgery
  - **Intervention**: prehabilitation



Joint work with Chet Hammill (Surgery), Jingwen Zhang, Dingwen Li, Ruixuan Dai (CSE)

Predict Complications after Pancreatic Surgery

Machine learning outperforms standard surgical risk scores.

- x2 AUPRC
- **x3 sensitivity at the same specificity**

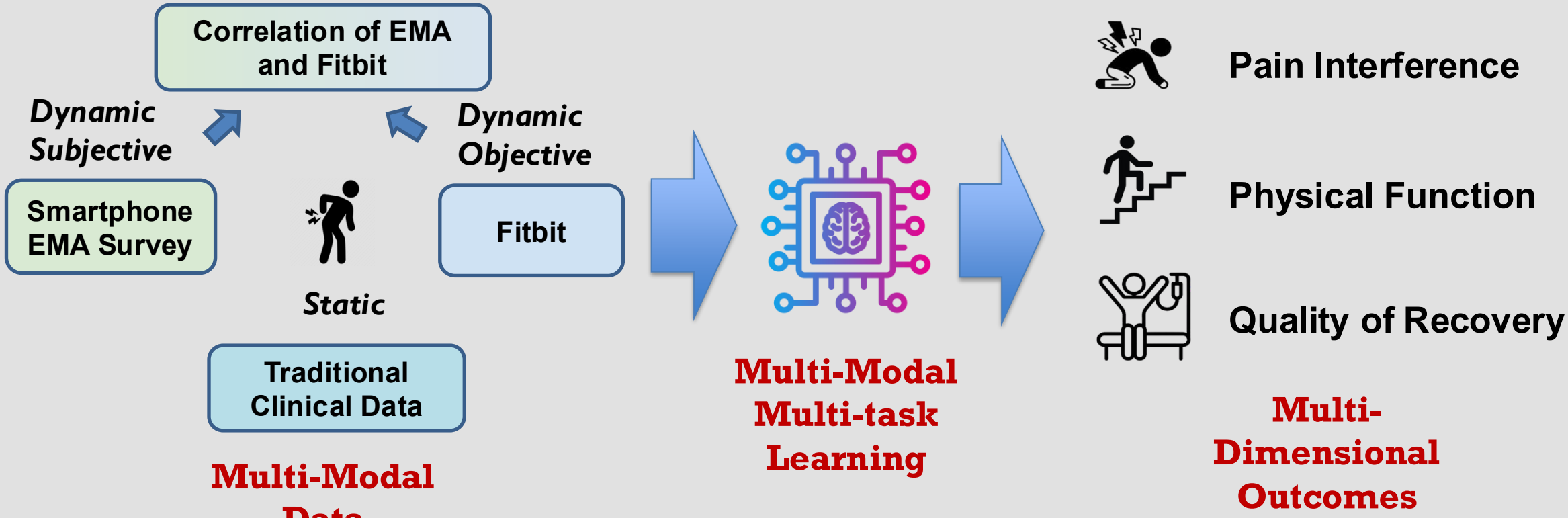
Data Source	AUROC	AUPRC	Sensitivity	Specificity
Random weighted classifier	0.5097 (0.0585)	0.4322 (0.0469)	0.1520 (0.0854)	0.8583 (0.0504)
NSQIP with Clinical Characteristics	0.6114 (0.0000)	0.4075 (0.0000)	0.2800 (0.0000)	0.8571 (0.0000)
ML with Clinical Characteristics	0.7632 (0.0085)	0.7374 (0.0206)	0.5800 (0.0699)	0.8583 (0.0083)
Wearable Data	0.7326 (0.0074)	0.7192 (0.0154)	0.5480 (0.0440)	0.8583 (0.0083)
Clinical Characteristics + Wearable Data	**0.8802 (0.0050)**	**0.8871 (0.0087)**	**0.8320 (0.0160)**	0.8583 (0.0083)

NSQIP: American College of Surgeons National Surgical Quality Improvement Program

Expanding to the Surgical Prehabilitation and Readiness (SPAR) program

J. Zhang, D. Li, R. Dai, H. Cos, G.A. Williams, L. Raper, C.W. Hammill, and C. Lu, Predicting Post-Operative Complications with Wearables: A Case Study with Patients Undergoing Pancreatic Surgery, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2), Article 87, July 2022.

Predicting Lumbar Spine Surgical Outcomes



Joint work with Jacob Greenberg (Neurosurgery), Ziqi Xu, Jingwen Zhang (CSE)

Xu Z, Zhang J, Greenberg J, Frumkin M, Javeed S, Zhang J, Benedict B, Botterbush K, Rodebaugh T, Ray W, Lu C. Predicting multi-dimensional surgical outcomes with multi-modal mobile sensing: A case study with patients undergoing lumbar spine surgery. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2024;8(2).

Predict Postoperative Risks from Notes

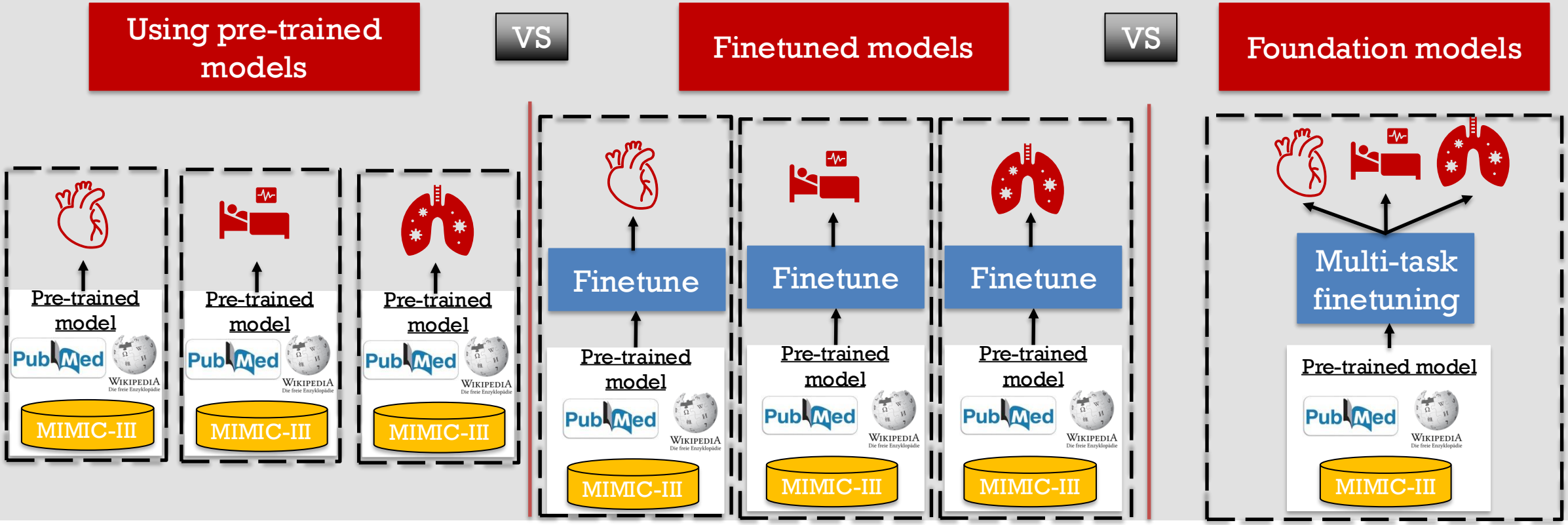
- Clinical notes holds valuable information about a patient.
- **Large language models (LLMs)** provide unprecedented capabilities to extract insights from texts.

- Data: 84,875 clinical notes from electronic anesthesia records at BJH
- Six outcomes
  1. 30-day mortality
  2. Acute knee injury (AKI)
  3. Pulmonary embolism (PE)
  4. Pneumonia
  5. Deep vein thrombosis (DVT)
  6. Delirium

Joint work with Charles Alba, Bing Xue (CSE), Joanna Abraham, Thomas Kannampallil (Anesthesiology)

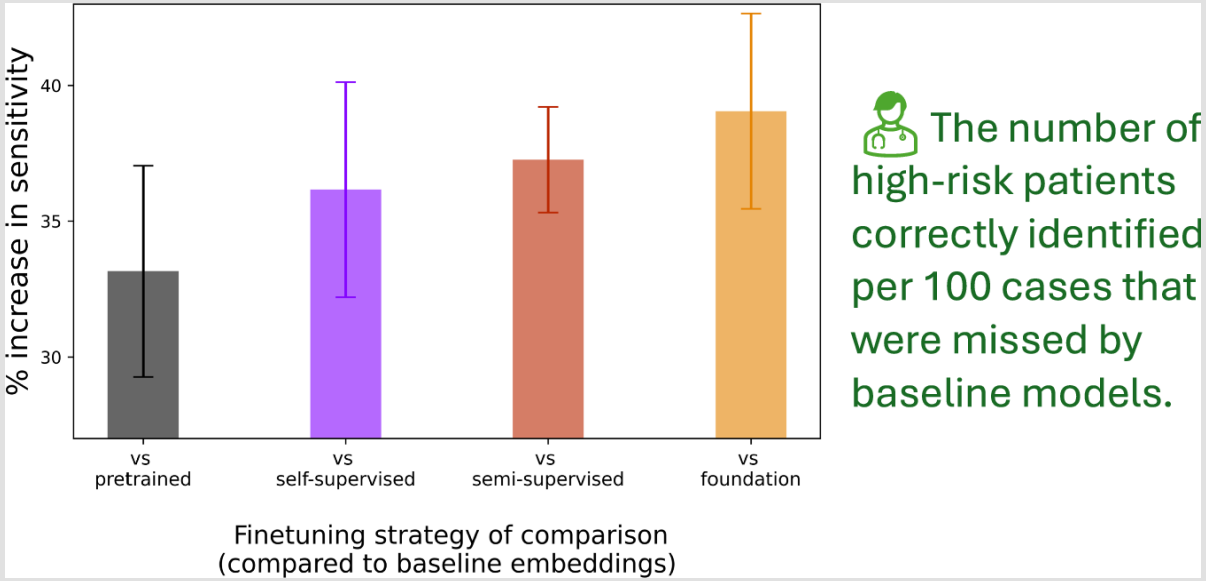
LLMs in Predicting Postoperative Risks

1. How does using pre-trained LLMs improve over traditional NLP approaches?
2. Can finetuning the models to surgical notes improve predictive performances?
3. Can a **foundation model** predict multiple risks?



LLMs in Predicting Postoperative Risks

- Pretrained LLMs hold promise in predicting postoperative risks
  - **33** percentage point increase in sensitivity over traditional NLP models (95% specificity)
- Foundation model tailored for surgical notes
  - Predict **multiple** postoperative risks.
  - **Outperform** LLMs fine-tuned for individual risks.
  - **Streamline** model integration into clinical workflows.

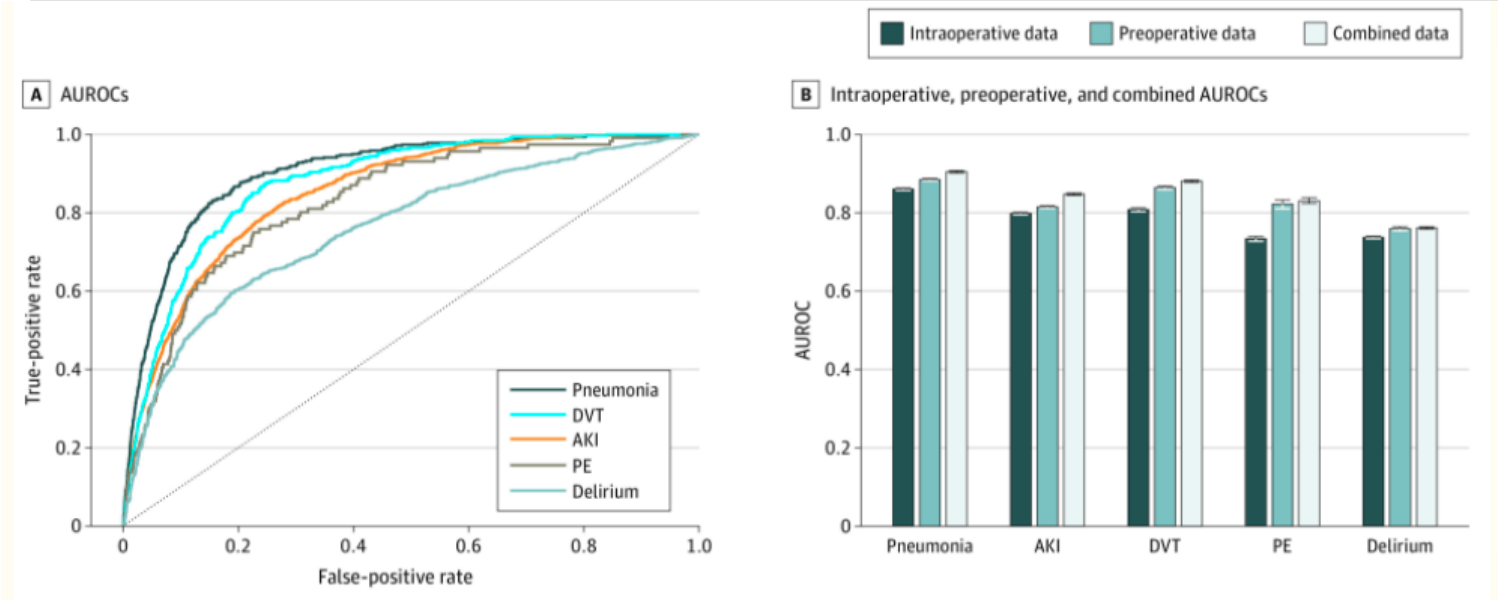


The number of high-risk patients correctly identified per 100 cases that were missed by baseline models.

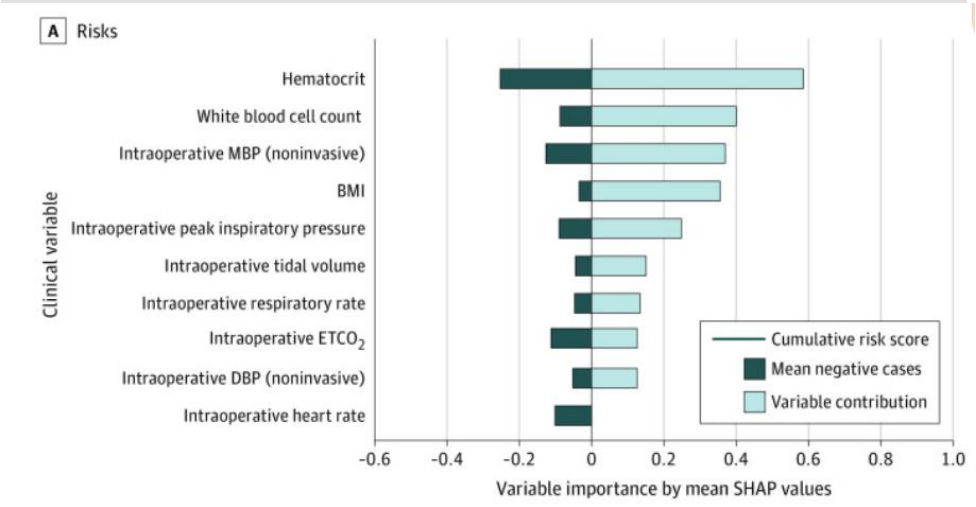
Alba C, Xue B, Abraham J, Kannampallil T, Lu C. The foundational capabilities of large language models in predicting postoperative risks using clinical notes. npj Digital Medicine. 8, 95 (2025).

Predict Postoperative Complications

- Anticipatory management for perioperative contingency planning
- Preemptive and early identification of risk factors using EHR data
- Example: OR-ICU handoff [ARHQ R01]



Predict multiple complications **Exploit pre- and intra-operative data**



Identify important risk factors

B. Xue, D. Li, C. Lu, C.R. King, T. Wildes, M.S. Avidan, T. Kannampallil, and J. Abraham, Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications, JAMA Network Open, 4(3):e212240, 2021.

Oncology Early Warning System

- **6.4%** of oncology inpatients have ICU transfer; **2.7%** of them die on the hospital wards
- Predict clinical deterioration based on heterogeneous EHR data
- **CrossNet**: Unified deep recurrent model for integrating static and time-series inputs

CrossNet predicts **10x** deterioration events than MEWS at the same false alarm rate.

MEWS: Modified Early Warning Scores

Model	Alarm rate control		False alarm control
	Sensitivity	Specificity	Sensitivity
MEWS	0.3358(0.0115)	0.8257(0.0142)	0.0392(0.0062)
C. BRITS	0.3899(0.0134)	0.9394(0.0097)	0.3581(0.0092)
S. BRITS	0.3891(0.0122)	0.9396(0.0105)	0.3586(0.0131)
CrossNet	**0.4218(0.0130)**	**0.9486(0.0093)**	**0.3952(0.0118)**

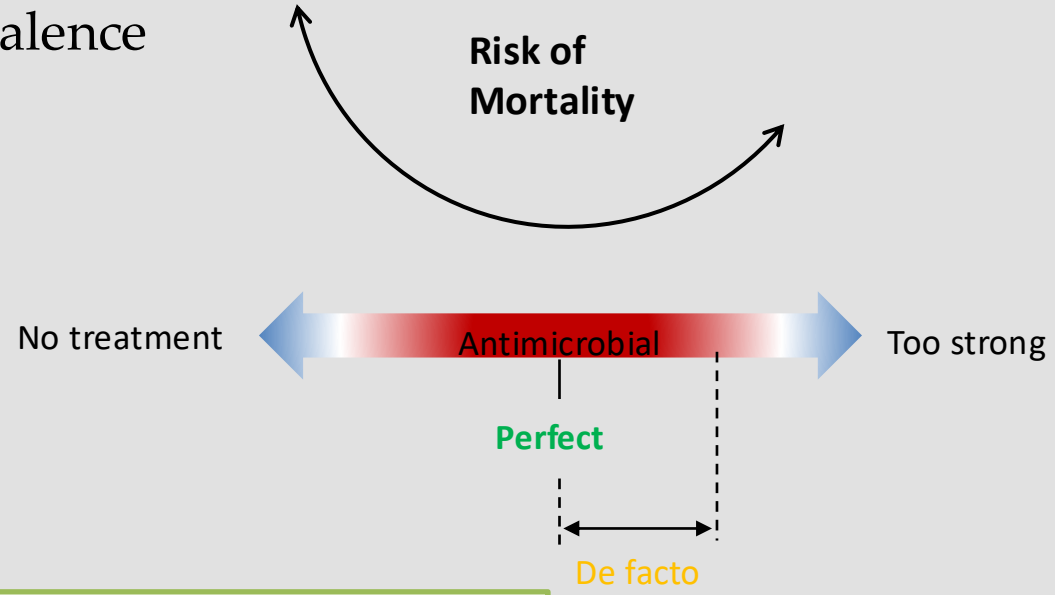
Li D, Lyons P, Klaus J, Gage B, Kollef M, Lu C. Integrating static and time-series data in deep recurrent models for oncology early warning systems. *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. 2021 Nov.

Predicting AMR in Sepsis Patients

- Developed a deep learning model that uses various EHR data (e.g., comorbidities, vitals, labs) to predict antimicrobial resistance (AMR) in sepsis patients

- Investigated performance across different hospitals and patient subgroups
  - Differences in patient case mix
  - Variations in local antimicrobial resistance prevalence

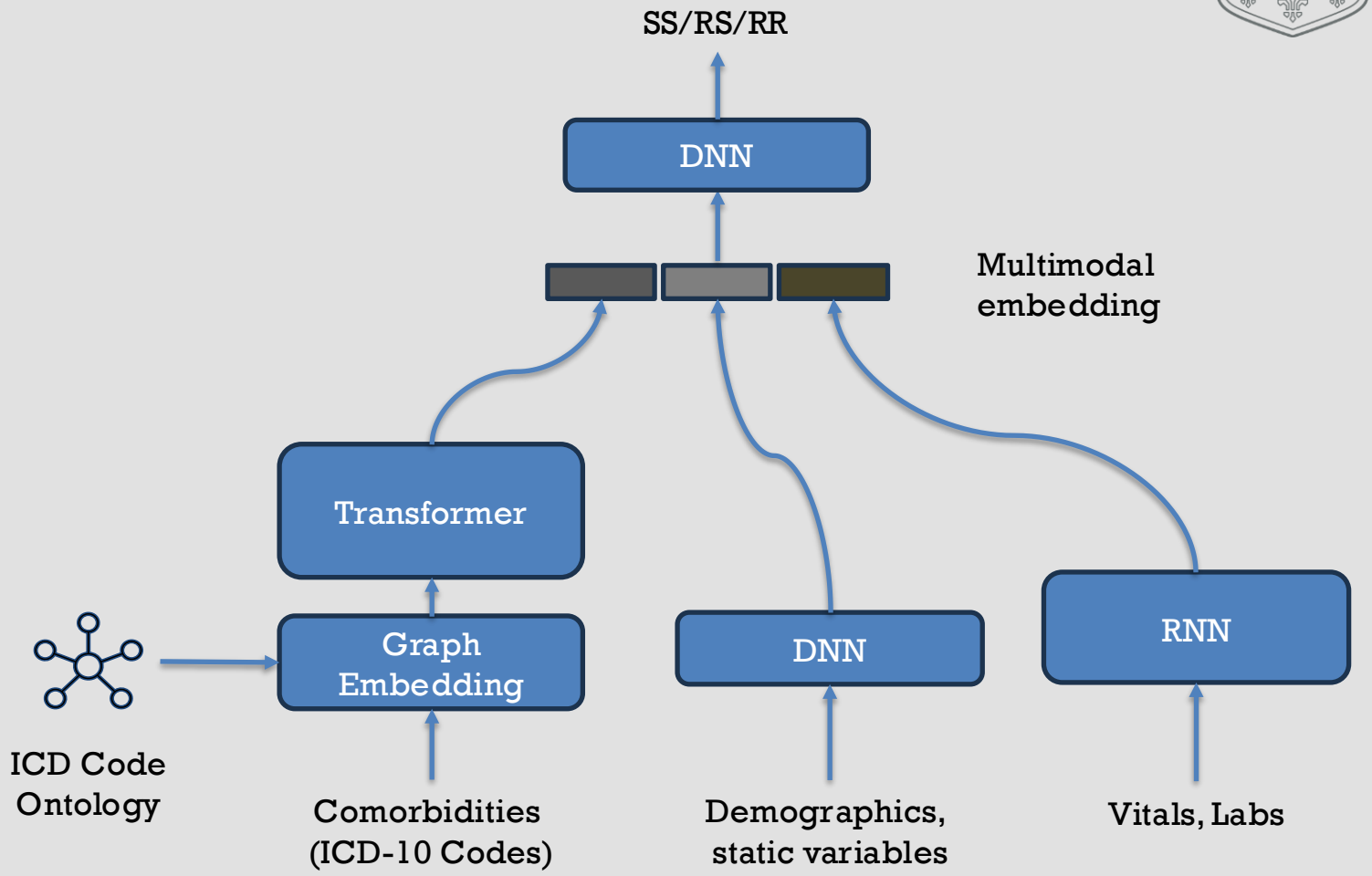
- Why does it matter?
  - Impacts on antibiotic stewardship
  - Sepsis mortality and treatment decisions
  - Policy and hospital-level decision making



Work with Vazquez Guillamet MC, Fraser VJ, Kollef MH (Medicine), Liu H (CSE)

Deep Learning Modeling Approach

- Data from 10 hospitals (2016–2021) with 39,893 patients and 85,238 sepsis episodes. Data components:
  - Demographics, static var.
  - Comorbidities
  - Vitals, labs
- Outcomes: Categorized gram-negative bacilli (GNB) resistance levels:
  - SS: Ceftriaxone-susceptible.
  - RS: Ceftriaxone-resistant, cefepime-susceptible.
  - RR: Resistant to both ceftriaxone and cefepime.

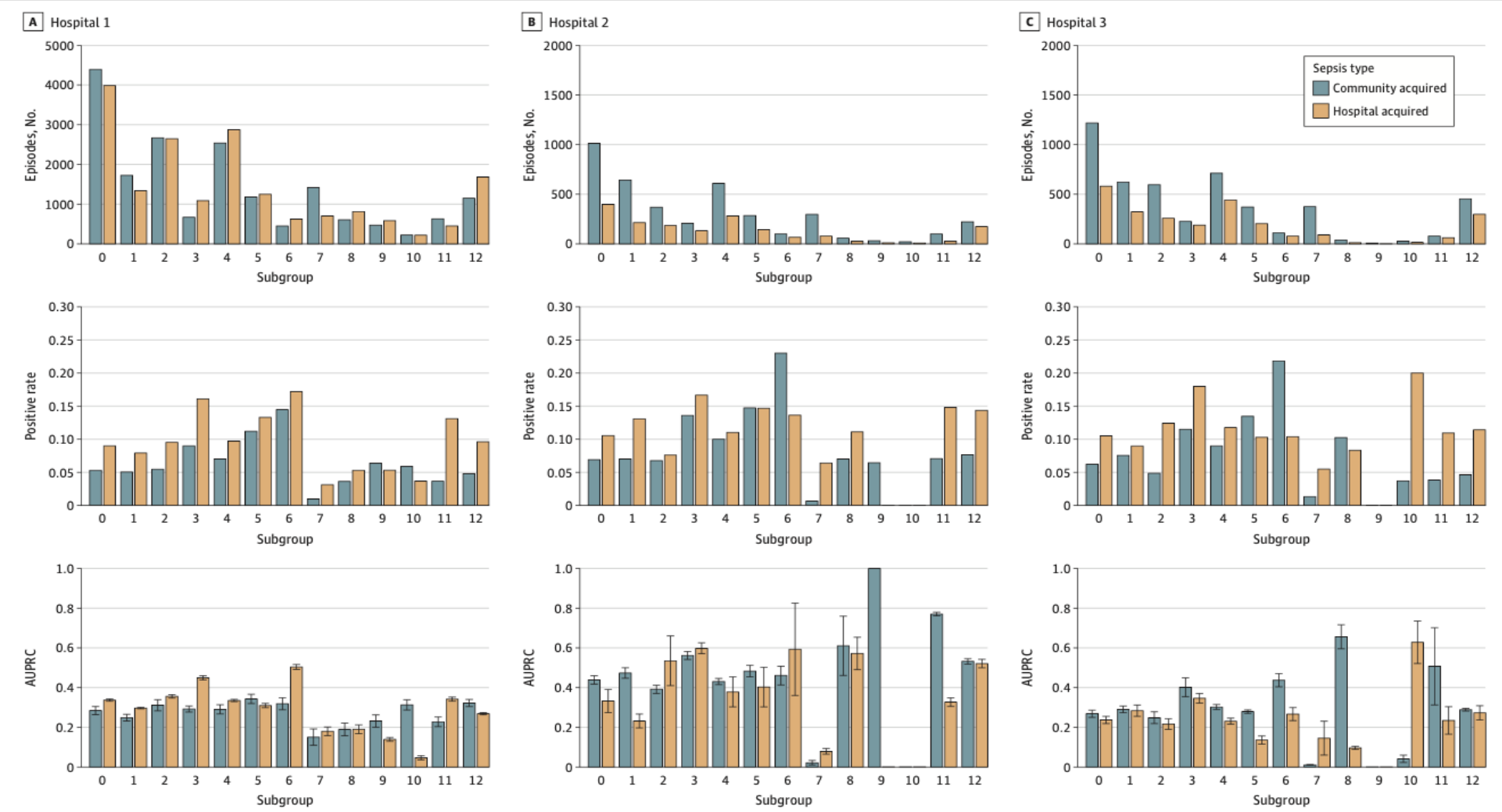


Performance:

- 0.84 AUROC (95%CI,0.83-0.84) in RS
- 0.81 AUROC (95%CI,0.79-0.83) in RR

Vazquez Guillamet MC, Liu H, Atkinson A, Fraser VJ, Lu C, Kollef MH. Performance of risk models for antimicrobial resistance in adult patients with sepsis. *JAMA Network Open.* 2024;7(11):e2443658.

Across Patient Case Mix and Hospitals



Barnes Baptist Med Ctr Christian

0: Entire cohort;
1: >=65 years;
2: <65 years;
3: History of bacterial pneumonia;
4: History of other sepsis;
5: History of disease-causing GNB such as E. coli, Klebsiella spp., and P. aeruginosa;
6: History of antibiotic-resistant microbes;
7: No history of 3, 4, 5, or 6;
8: Hematological malignancies;
9: Transplantation;
10: Alcoholic cirrhosis;
11: Septic shock;
12: Intubation.

Predict Treatment Adherence

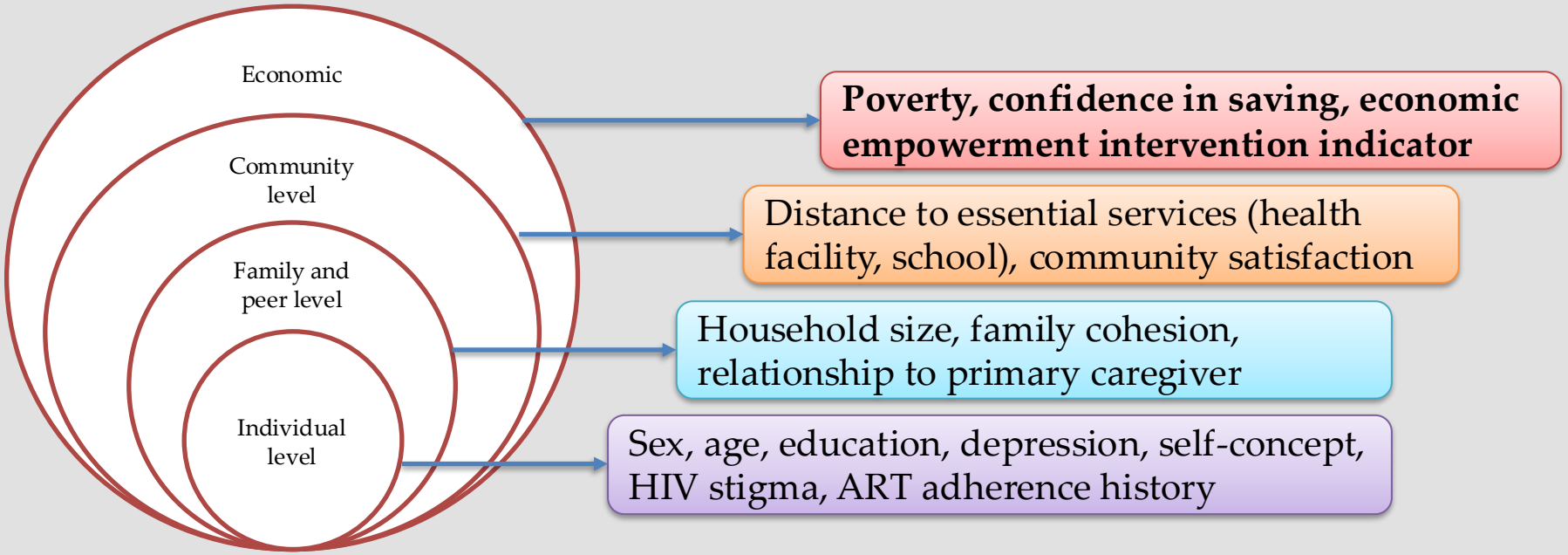
- Approx. **85%** of 1.7 million adolescents living with HIV (ALWHIV) live in SSA
- Globally, viral suppression among adolescents remains low
  - 72% viral suppression among adults with HIV
  - **46%** viral suppression among ALWHIV
- Current HIV treatment monitoring methods only allow for reactive action
- Early identification of ALWHIV at risk for non-adherence is crucial
- **Goal:** Predict poor adherence to antiretroviral therapy in ALWHIV

Work with Claire Najjuuko (DCDS), Ziqi Xu (CSE), Rachel Brathwaite, Samuel Kizito, Fred Ssewamala (ICHAD)

Predictors and Outcome

Predictors: **Multi-level socio-behavioral** risk factors

Outcome: **ART adherence** measured using self-reports and viral load results

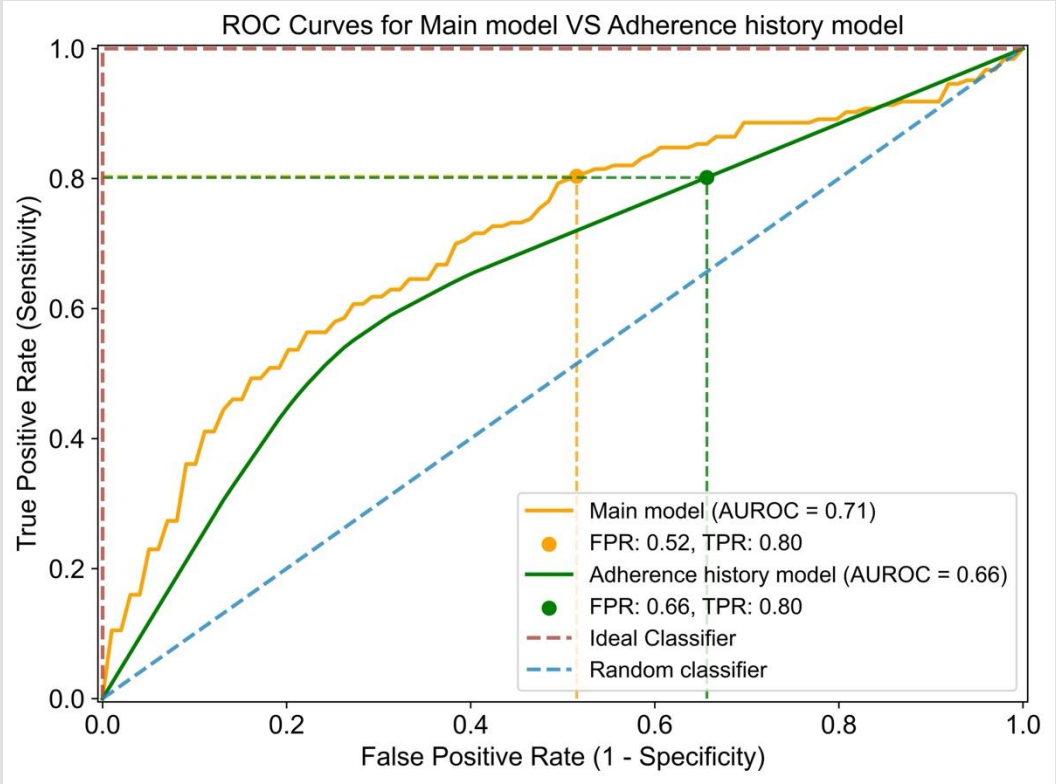


Predictive Performance

Model	AUROC	AUPRC	Sensitivity	Specificity
Random forest	0·71 (0.07)	0·55 (0.09)	0·80 (0.09)	0·48 (0.00)
Logistic regression	0·70 (0.07)	0·53 (0.07)	0·81 (0.09)	0·41 (0.00)
Lasso	0·70 (0.07)	0·53 (0.07)	0·81 (0.09)	0·41 (0.00)
Support vector machine	0·69 (0.05)	0·54 (0.08)	0·81 (0.06)	0·35 (0.00)
Gradient boosting	0·69 (0.06)	0·54 (0.07)	0·80 (0.11)	0·39 (0.00)
xGBoost	0·68 (0.07)	0·51 (0.09)	0·81 (0.12)	0·39 (0.00)
Decision trees	0.61 (0.03)	0.40 (0.04)	0.80 (0.03)	0.24 (0.00)

Adding socio-behavioral and economic factors cut the false alarm rate by **14 points** compared to a history-only model, at the same sensitivity.

Machine learning enables **predicting future risk** of poor adherence.



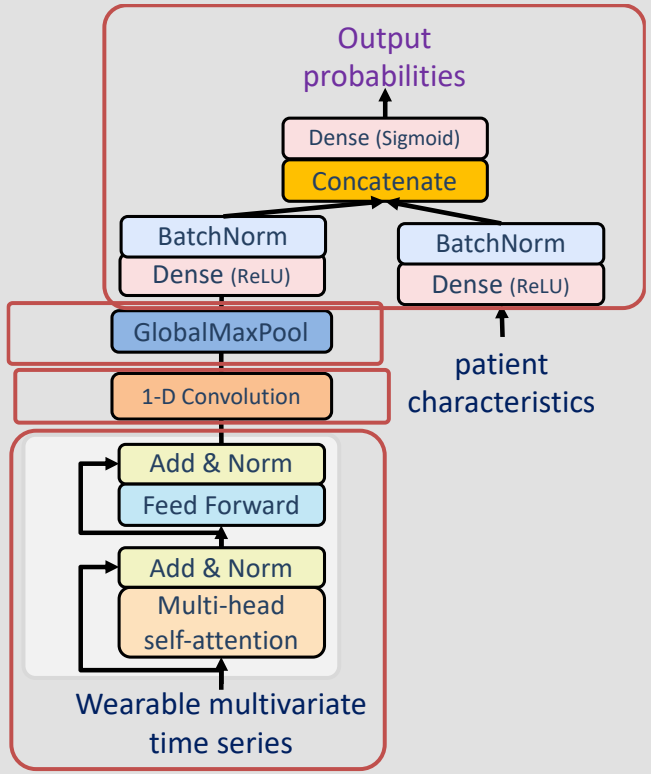
Najjuuko C, Xu Z, Kizito S, Lu C, Ssewamala FM. Using machine learning to predict poor adherence to antiretroviral therapy among adolescents living with HIV in low-resource settings. AIDS. 2025;39(9):1204–1213.

Important Predictors

- **History of adherence** is the most important predictor for future adherence
- **Economic factors** are highly associated with future non-adherence

Detect Depression and Anxiety Disorders

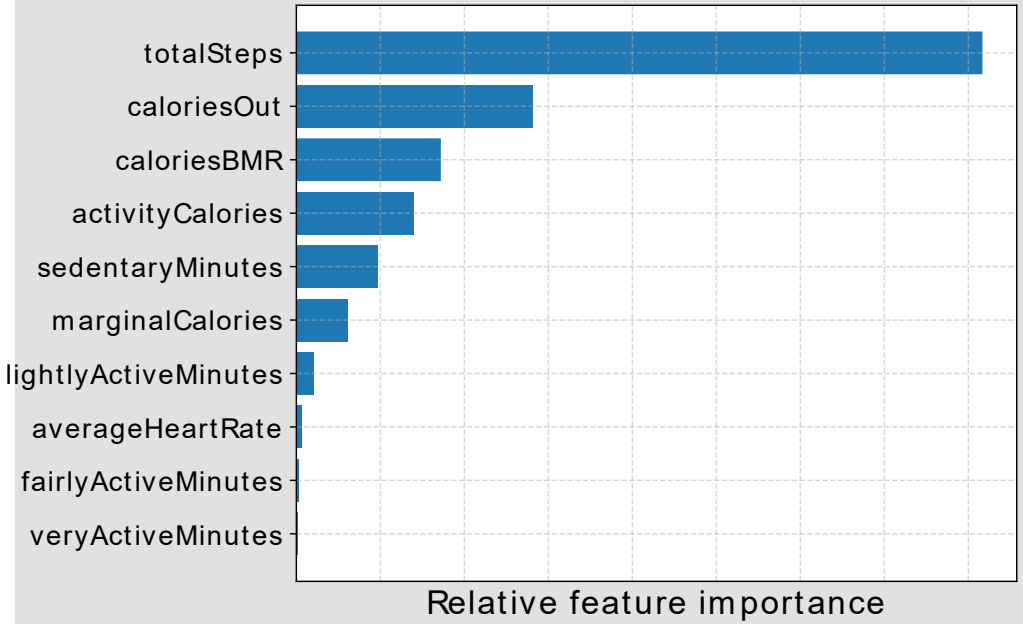
- **3.8%** of the population (**280 million**) experience depression (WHO).
- Over **50%** of patients are not recognized or treated.
- Clinical visit is time-consuming and expensive.
- **WearNet**: Deep learning for detecting depression and anxiety disorders
  - **Unobtrusive** screening with wearables
  - Enable **timely**, **targeted** intervention
  - Trained on NIH All of Us: 8,996 participants (1,247 positive)

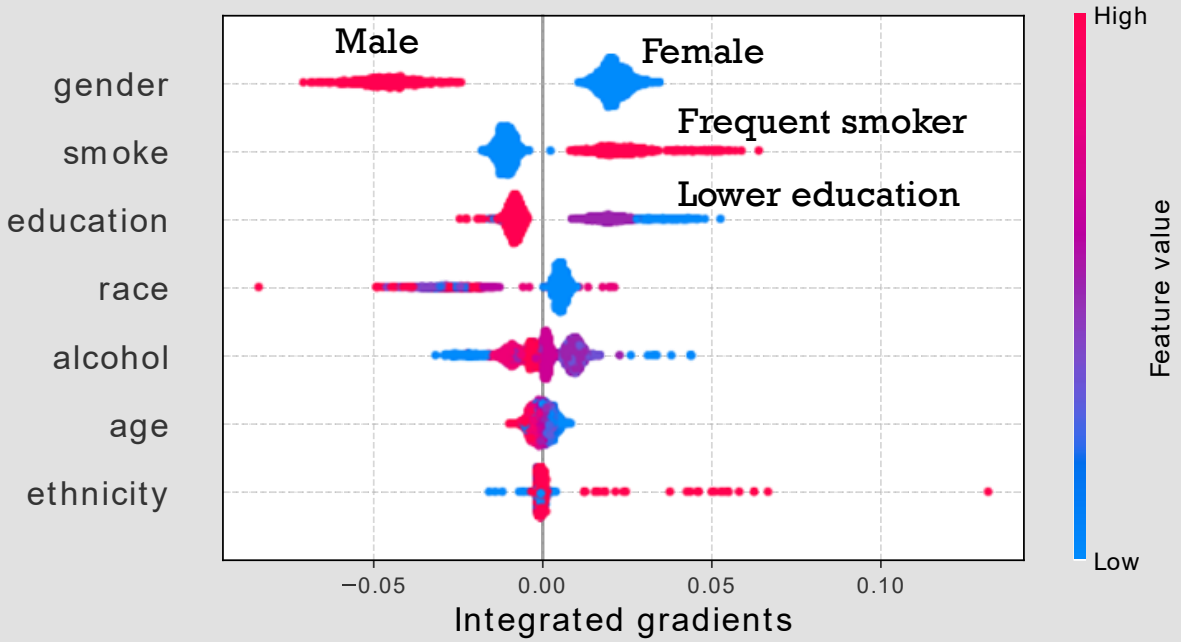


Dai R, Kannampallil T, Kim S, Thornton V, Bierut L, Lu C. Detecting mental disorders with wearables: A large cohort study. *Proceedings of the ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI)*. 2023 May.

Model Explanation

- Assign importance scores to features by approximating the integral of gradients





- **Total steps** is the most important wearable feature.
- **Women** and **frequent smokers** are more likely to be diagnosed.

Match literature

Phenotype Chiari Malformation Type 1 (CM1)

- CM1 is a **heterogeneous** disease with significant variability in clinical presentation and natural history.
- There is growing recognition that CM1 likely represents a constellation of clinical phenotypes rather than a singular disease entity.
- Goal: Identify **distinct, clinically meaningful** phenotypes

Park Reeves Syringomyelia Research Consortium registry: 1340 patients from 42 medical centers since 2011

Orthopedic Info
e.g., Spinal Fusion

General Exam
e.g., Head Circumference

Symptoms
e.g., Neck Pain

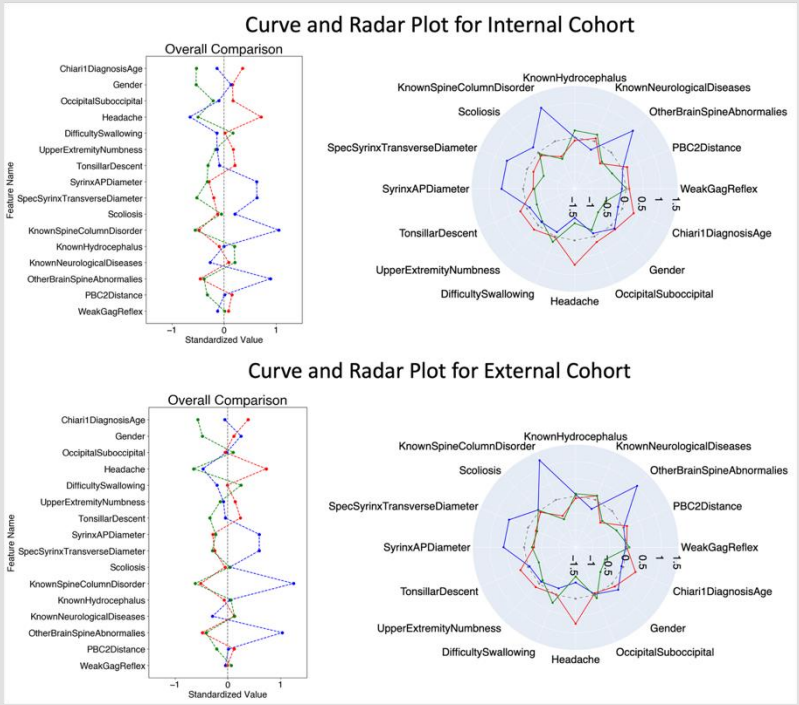
Input Features
(500+ variables)

Family History
e.g., Parent Syringomyelia

Childhood History
e.g., Chronic Headache

Radiographic Exam
e.g., Syrinx AP Diameter

Neurological Exam
e.g., Facial Weakness



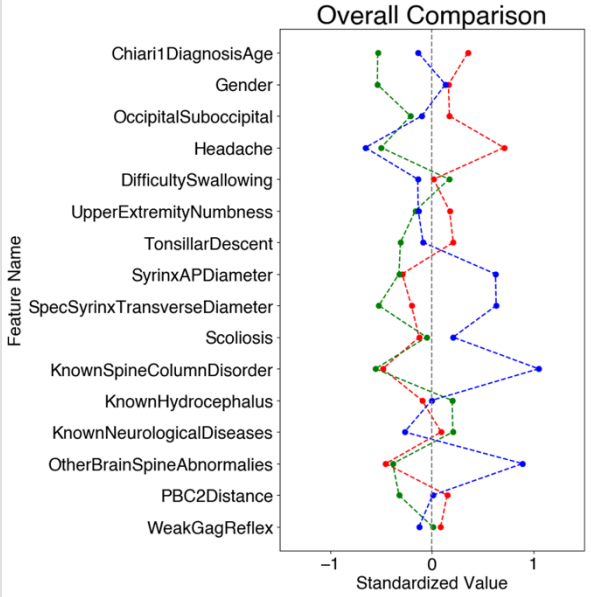
Work with Dave Limbrick, Jacob Greenberg, Sean Gupta (Neurosurgery), Ziqi Xu (CSE) (NIH P01)

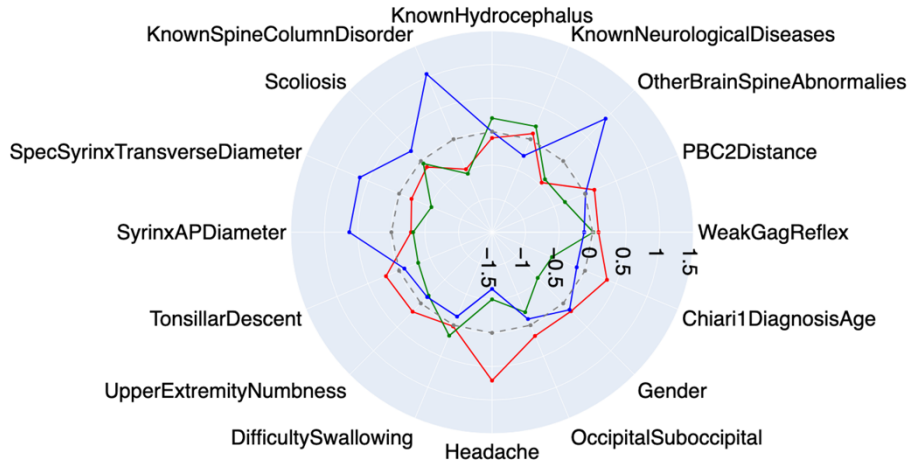
Clustering

- **Identified three distinct clusters**
- **Reproducible across internal/external cohorts**
- **Inform clinical decision-making and management**

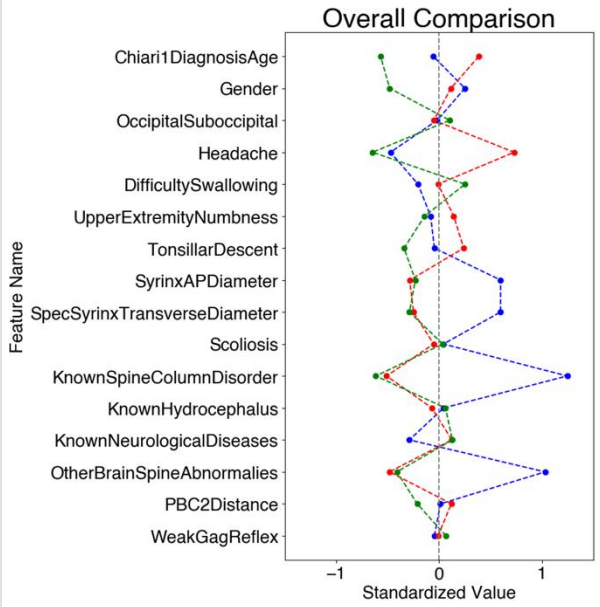
Gupta VP, Xu Z, Greenberg JK, Strahle JM, Haller G, Meehan T, Roberts A, Limbrick DD Jr, Lu C. Using artificial intelligence to identify three presenting phenotypes of Chiari type-1 malformation and syringomyelia. Neurosurgery. 2024 Nov 18.

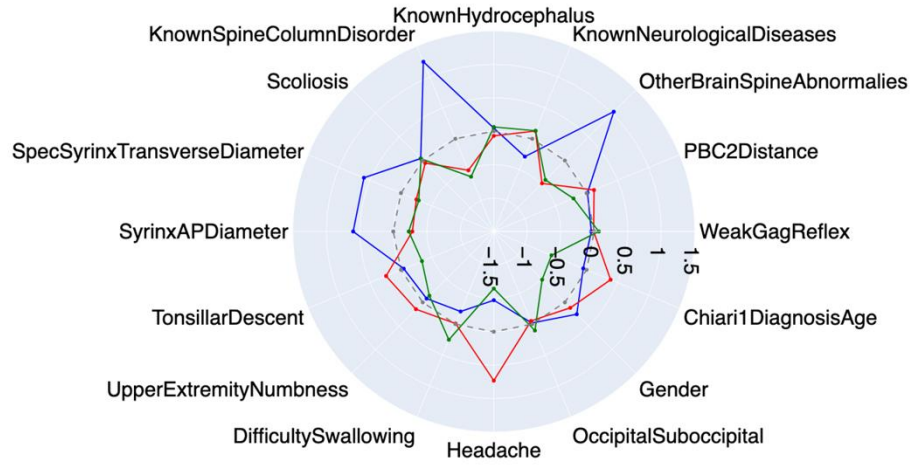
Curve and Radar Plot for Internal Cohort





Curve and Radar Plot for External Cohort





Advancing Health with AI

- AI as a transformative tool for health care
  - **Predicting** outcomes → Personalized care strategies
  - Model **interpretation** → Actionable health insights
  - **Phenotyping** diseases and population health → Precision interventions

- Extract insights from **multi-modal, high dimensional, noisy** data

- **Integration** of data, health expertise, and AI is essential.

AI for Health Institute

- **Hub** for interdisciplinary innovation at the intersection of AI and health
- **Catalyst** for collaborative teams across domains
- **Magnet** for top talent in AI and health sciences
- **Leader** of large-scale, high-impact initiatives
- **Accelerator** for translating AI into clinical and public health practice

Bringing AI and Clinical Experts Together to Tackle Critical Health Challenges

https://aihealth.wustl.edu/

Launched at AI & Digital Health Summit, 10/18/2023

AIHealth Community

- **100+** faculty members across **7** schools and **26** departments/divisions

- Steering committee: Michael Avidan (Anesthesiology), Aaron Bobick (Engineering), Lisa Connor (Occupational Therapy), Richard Cote (Pathology & Immunology), Vicky Fraser (Medicine), Roch Guerin (CS), Eric Lenze (Psychiatry), John Lynch (BJC), Thomas Maddox (BJC), Bill Powderly (ICTS), Philip Payne (Informatics), Pamela Woodard (Radiology), Gregory Zipfel (Neurosurgery)

AI for Health Institute

Driving Innovations at the Intersection of AI and Health

- AIHealth **Seed Funding** Program
- AIHealth **Symposium**
- **Connect** AI and clinical experts

AI
- Equity & fairness
- Human-in-the-loop
- Imaging AI
- Large language models
- Security & trust
- Wearables & mobile

Health
- Cancer care
- Digital pathology
- Global health
- Infectious diseases
- Medical imaging
- Mental health
- Perioperative care
- Public health
- Reproductive health

https://aihealth.wustl.edu/

AI for Health Symposium

Can We Harness AI to Promote Healthier Lives?

- Hosts: School of Public Health, McKelvey School of Engineering, AI for Health Institute
- Wednesday, October 22
- Whitaker 100 and online
- https://aihealth.wustl.edu/symposium/