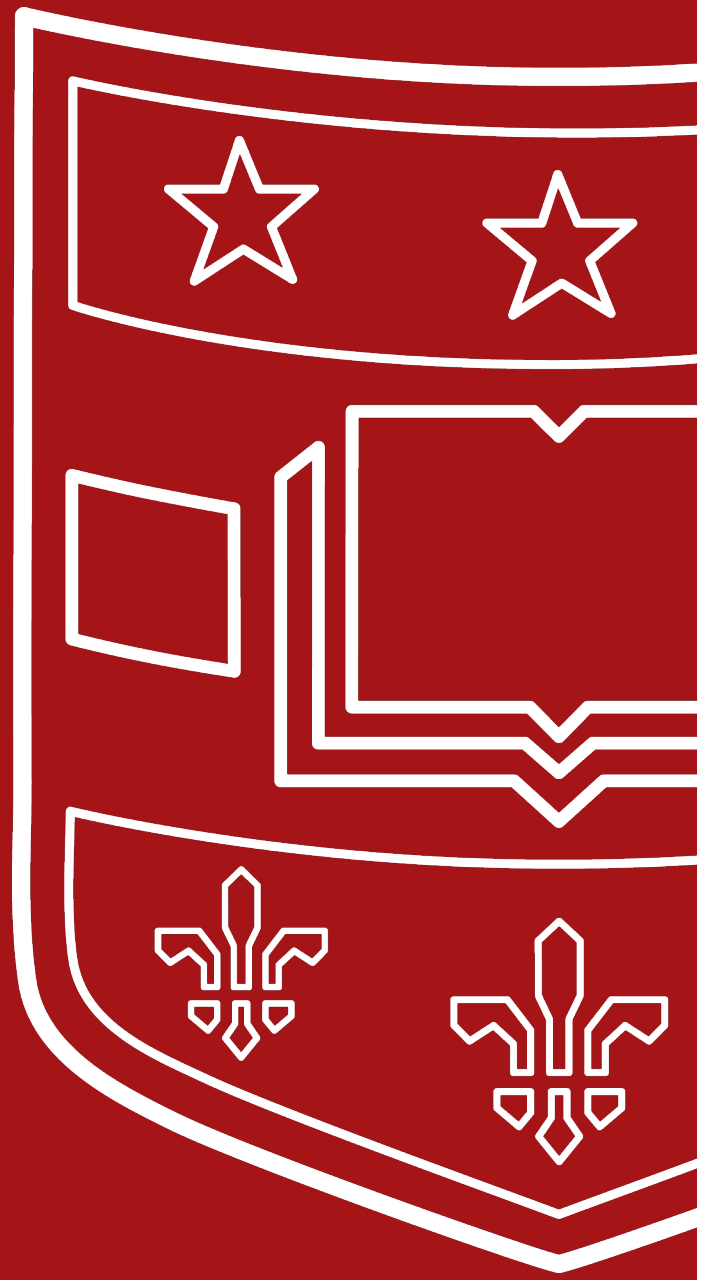
CSE 4109

Introduction to AI for Health

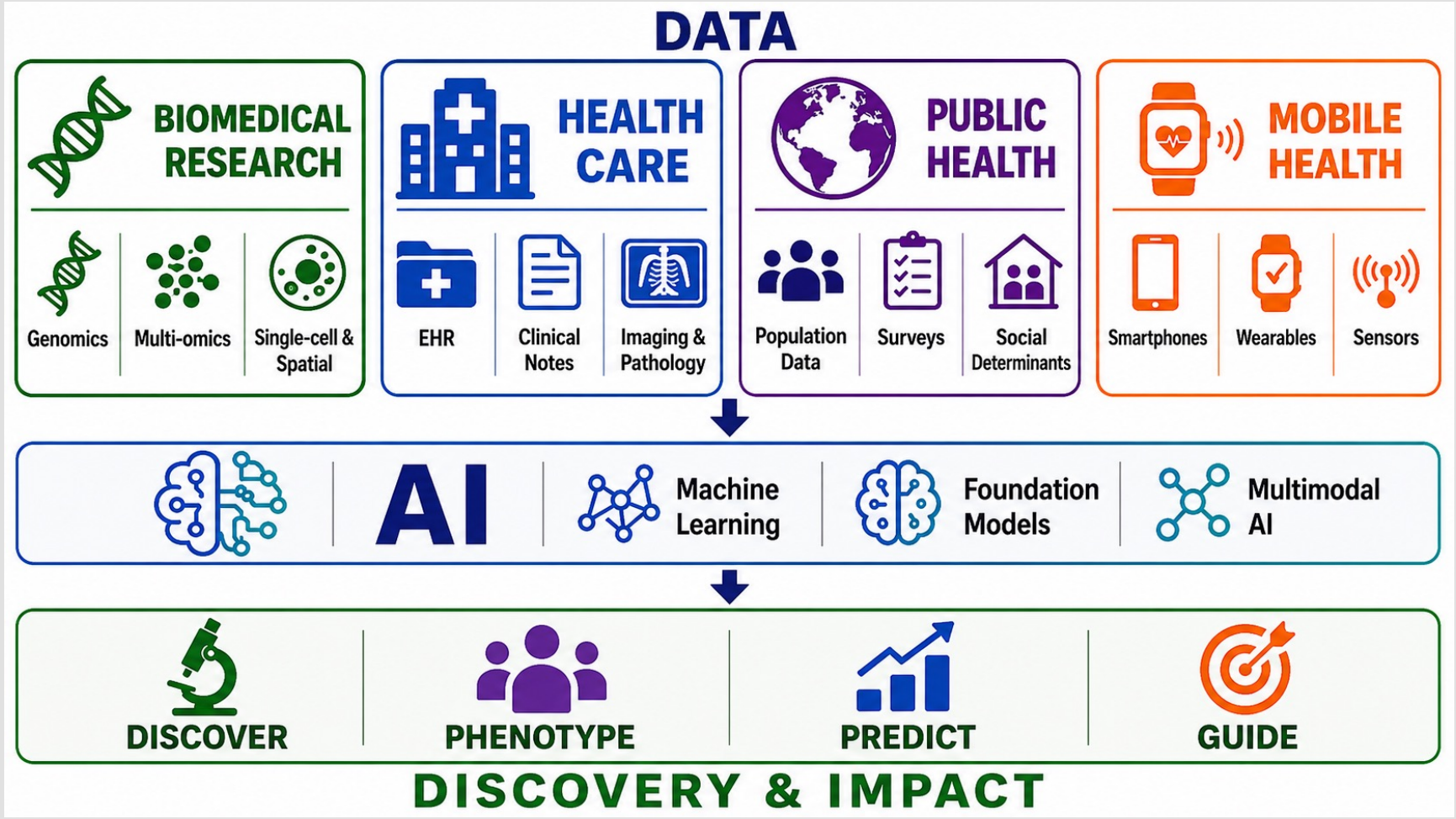
Chenyang Lu
Director, AI for Health Institute (AIHealth)
Fullgraf Professor
Computer Science · Anesthesiology · Medicine · Neurosurgery · Psychiatry · Public Health



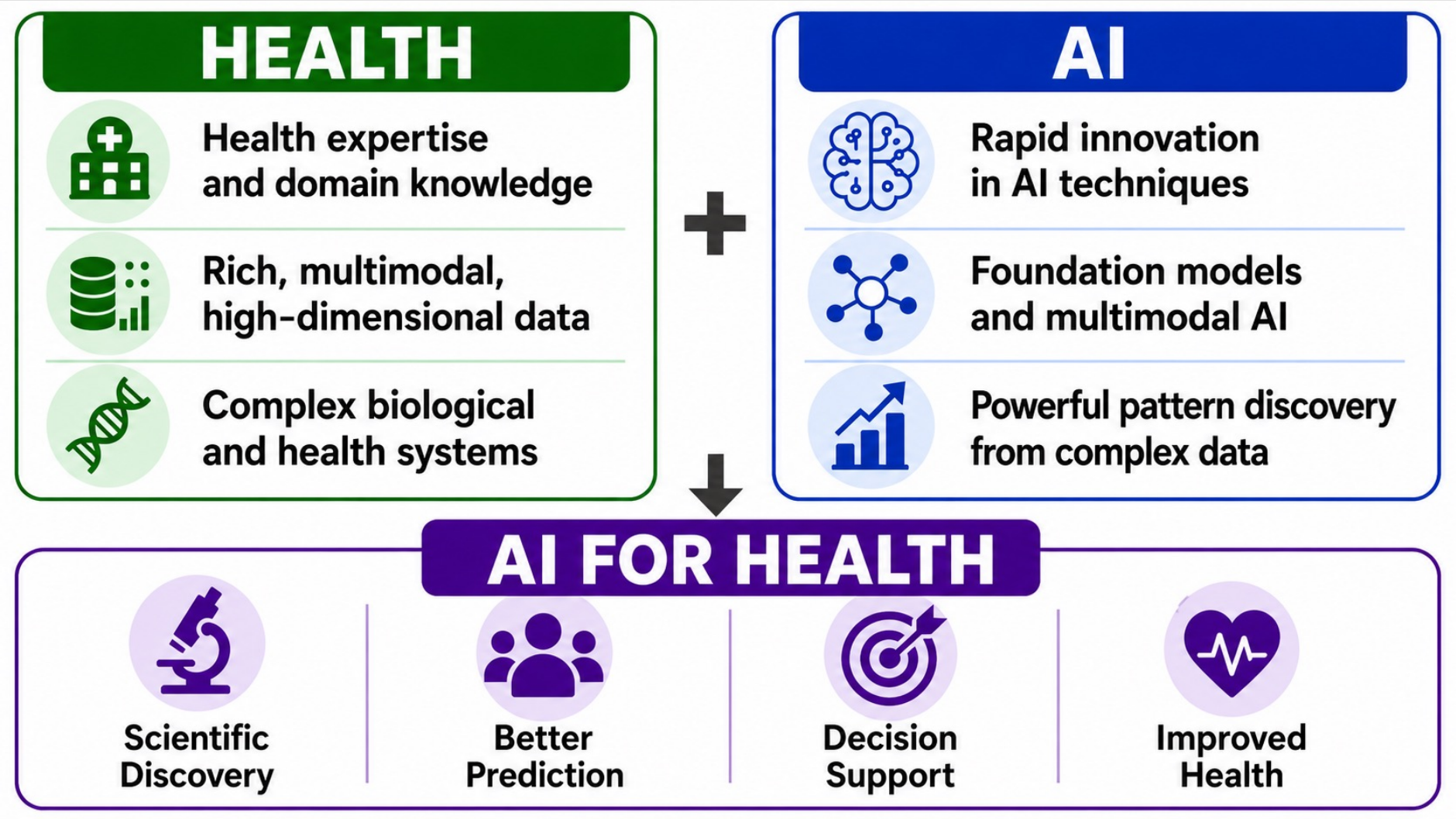
Washington University in St. Louis

AI for Health: Data → Impact

- **Discover** disease mechanisms and therapeutic targets
- **Phenotype** diseases, populations, and biological systems
- **Predict** outcomes and treatment responses
- **Guide** clinical and public health decisions



Health + AI = AI for Health



Three Frontiers in AI for Health

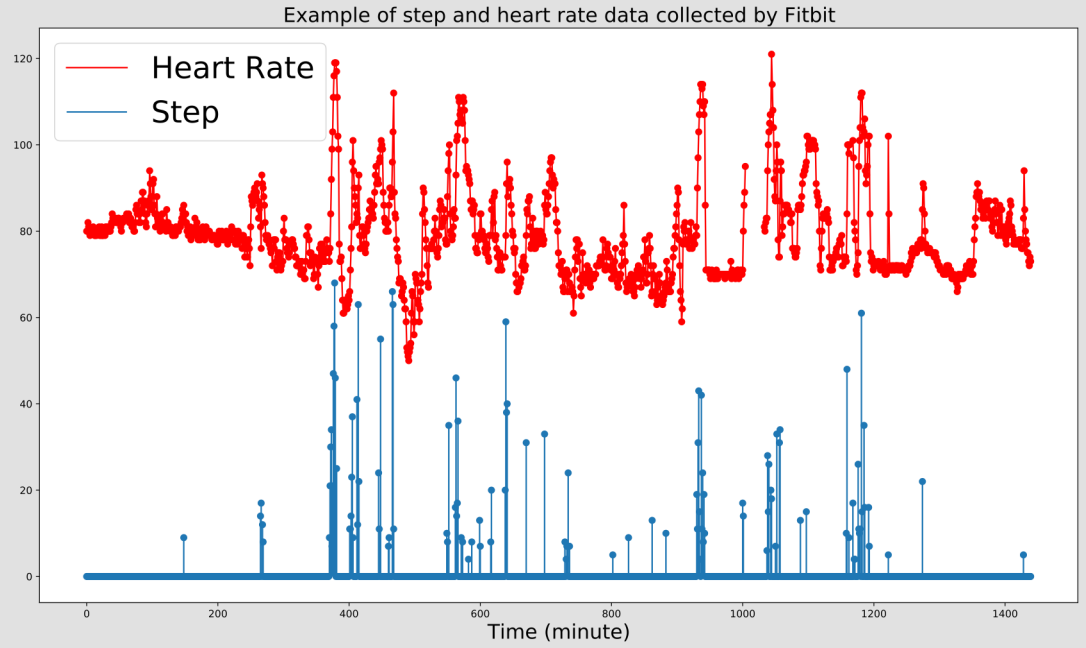
- Predicting clinical outcomes using wearables
- Phenotyping complex diseases and health service utilization
- Foundation models for multimodal health data

Three Frontiers in AI for Health

- **Predicting clinical outcomes using wearables**
- Phenotyping complex diseases and health service utilization
- Foundation models for multimodal health data

Wearables

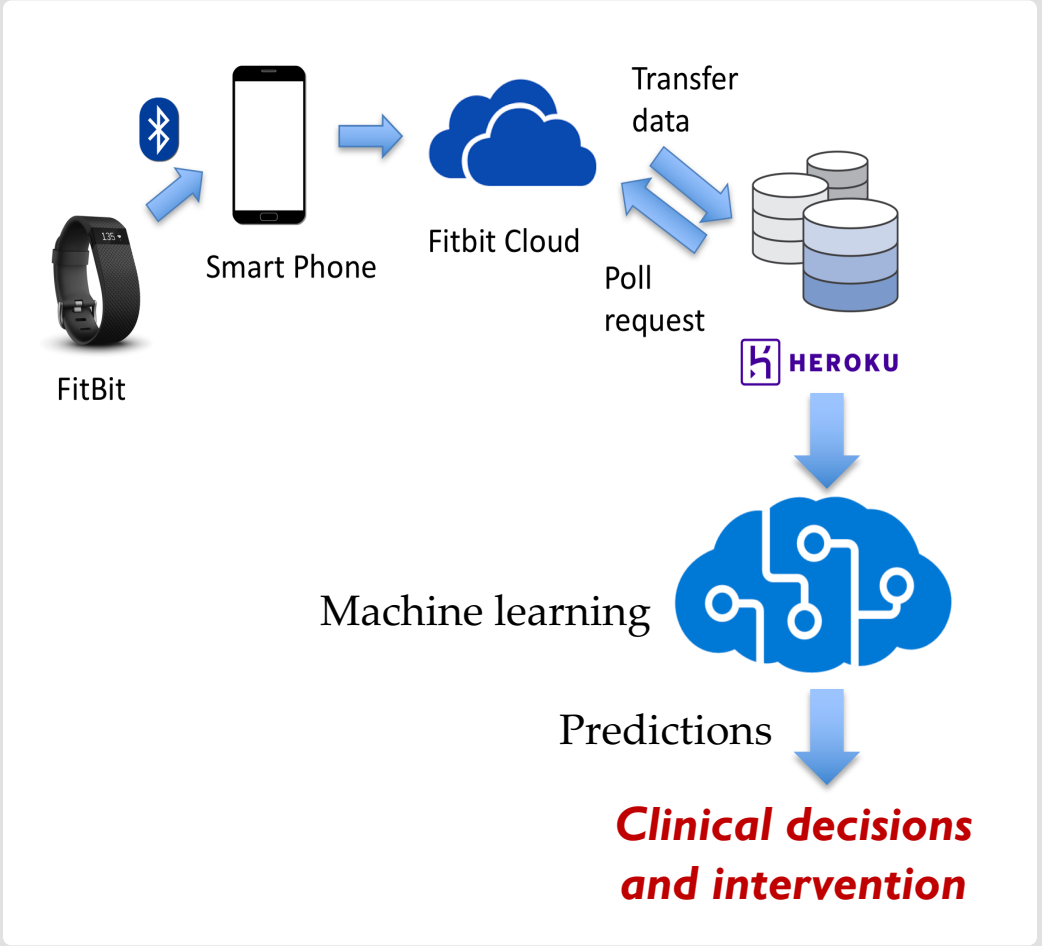
- Commonly available: step, heart rate, sleep stages
- More sensing modalities
  - Oxygen saturation (SpO2)
  - Skin temperature
  - Breathing rate
  - Heart rate variability
  - ECG
  - Stress
- **500+ million** wearables sold in 2021



Unobtrusive, longitudinal monitoring outside hospitals!

Decision Support for Pancreatic Surgery

- Pancreatic cancer has a 5-year survival rate less than 5%.
- Surgery is 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



Work with Chet Hammill (Surgery), Jingwen Zhang (CSE)

Predict Complications after Pancreatic Surgery

Machine learning outperforms standard surgical risk scores.

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

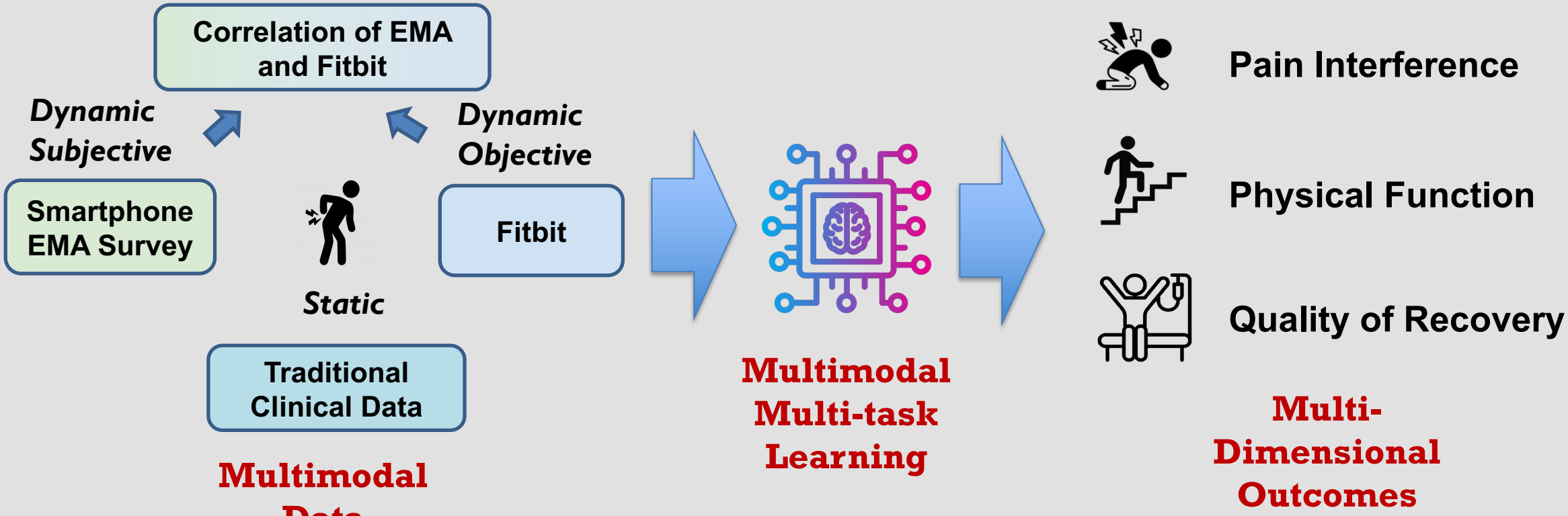
Data Source	AUROC	AUPRC	Sensitivity	Specificity
Random weighted classifier	0.5097 (0.0585)	0.4322 (0.0469)	0.1520 (0.0854)	0.8583 (0.0504)
NSQIP with Clinical Characteristics	0.6114 (0.0000)	0.4075 (0.0000)	0.2800 (0.0000)	0.8571 (0.0000)
ML with Clinical Characteristics	0.7632 (0.0085)	0.7374 (0.0206)	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.

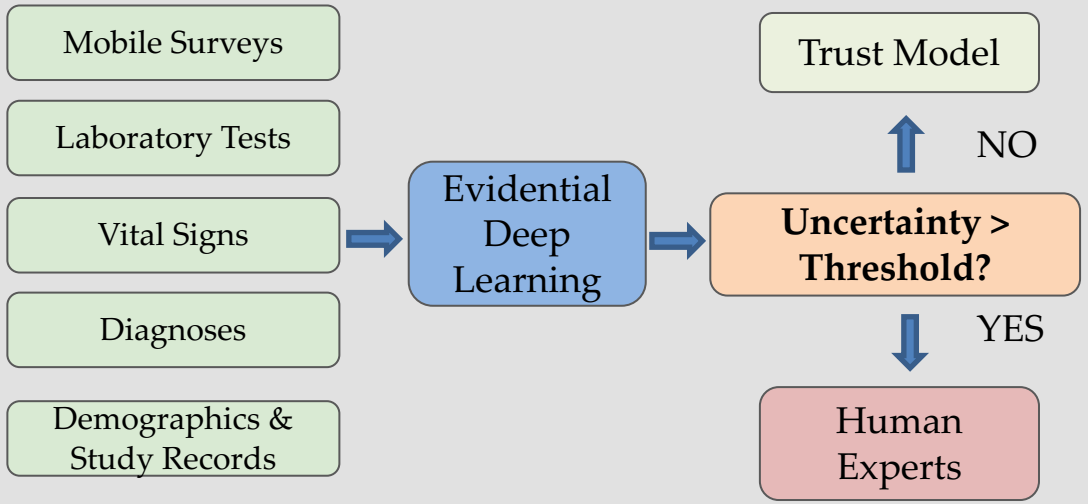
Predicting Lumbar Spine Surgical Outcomes

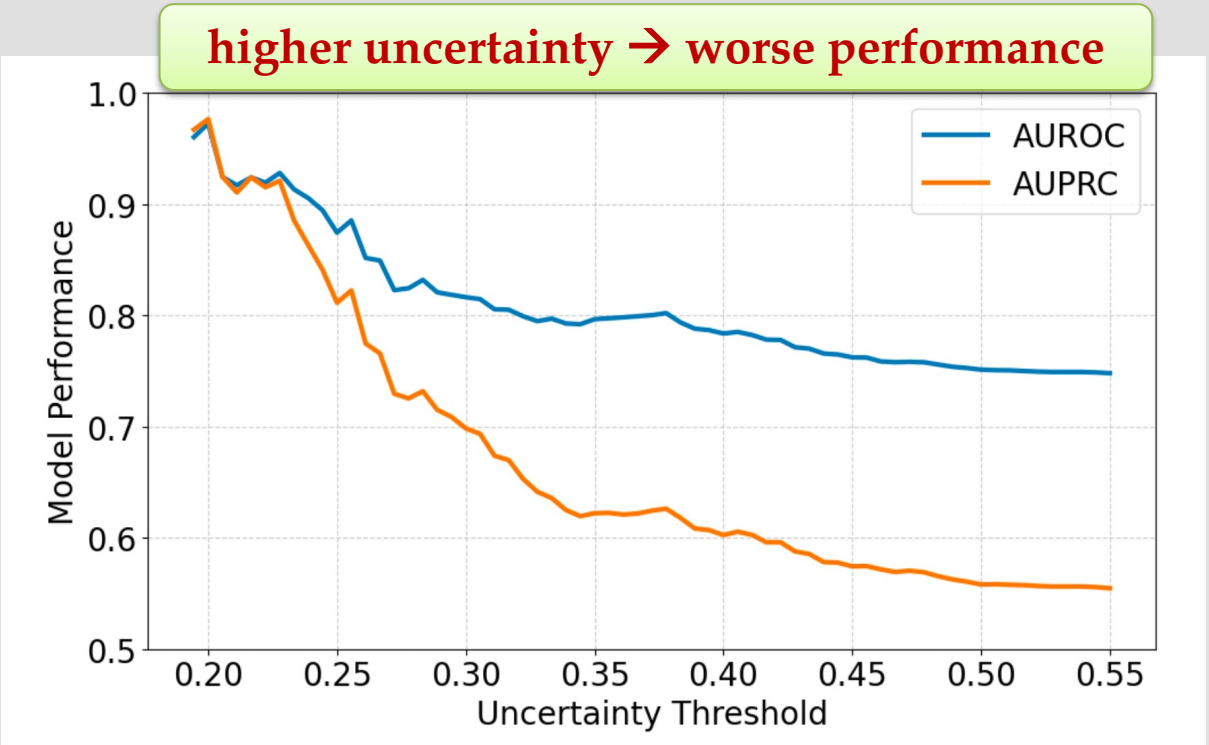


Xu Z, Zhang J, Greenberg J, Frumkin M, Javeed S, Zhang J, Benedict B, Botterbush K, Rodebaugh T, Ray W, Lu C. Predicting multi-dimensional surgical outcomes with multi-modal mobile sensing: A case study with patients undergoing lumbar spine surgery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.* 2024;8(2).

Predict Postsurgical Persistent Pain

- Pain persists 3+ months after surgery
- Multimodal data: pre-operative mobile sensing and clinical data
- **Uncertainty-aware**: support clinical decisions with estimated uncertainty

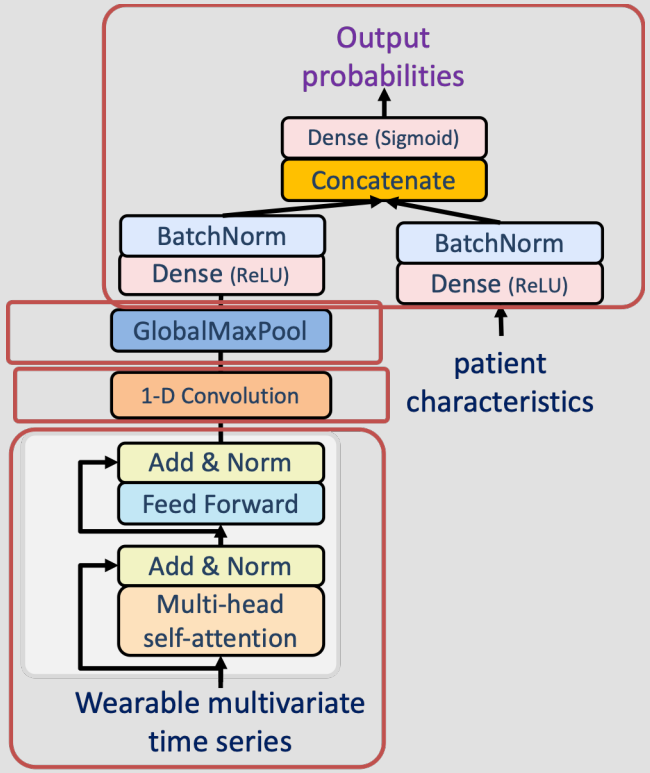




Xu Z, Zhang J, Haroutounian S, Liu H, Cao Z, Messner GR, Alaverdyan HB, Ahuja S, Koshy R, Hanns J, Frumkin M, Rodebaugh TL, Lu C. Incorporating uncertainty in predictive models using mobile sensing and clinical data: a case study on persistent post-surgical pain. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2025;9(2):58.

Detect Depression and Anxiety Disorders

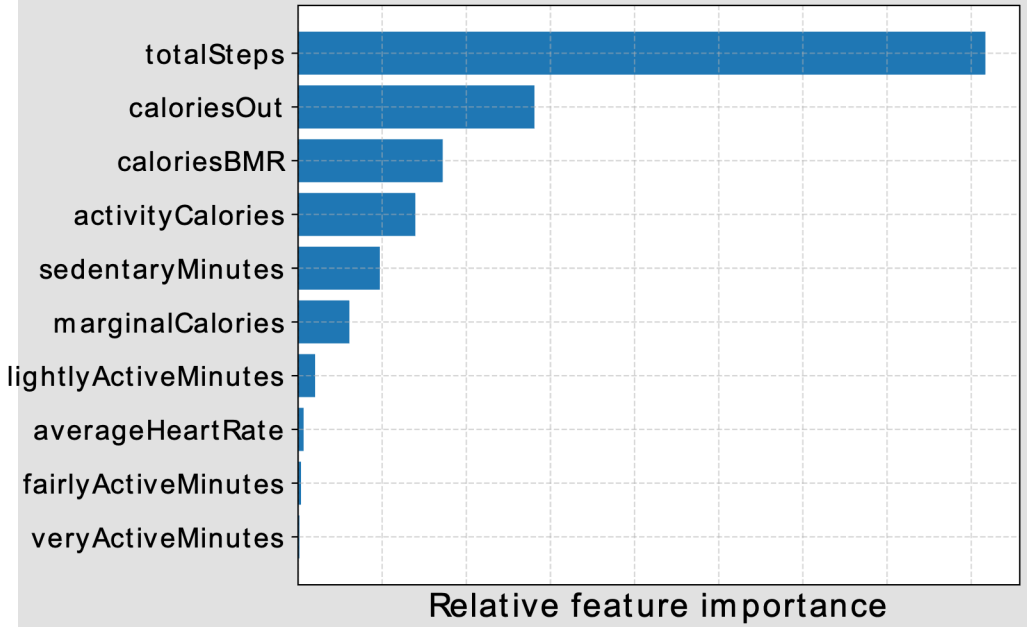
- **3.8%** of the population (**280 million**) experience depression (WHO).
- Over **50%** of patients are not recognized or treated.
- Clinical visit is time-consuming and expensive.
- **WearNet**: Detect depression and anxiety disorders
  - **Unobtrusive** screening with wearables
  - Enable **timely**,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

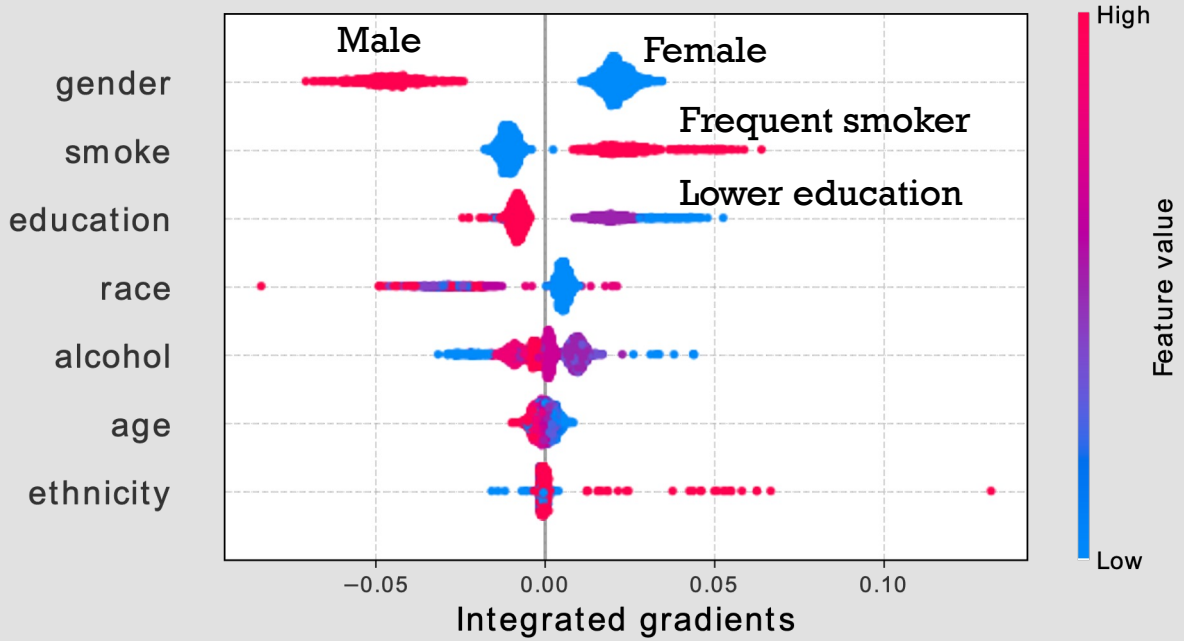


Work with Thomas Kannampallil (Anesthesiology), Laura Bierut (Psychiatry), Ruixuan Dai (CSE)

Model Explanation

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





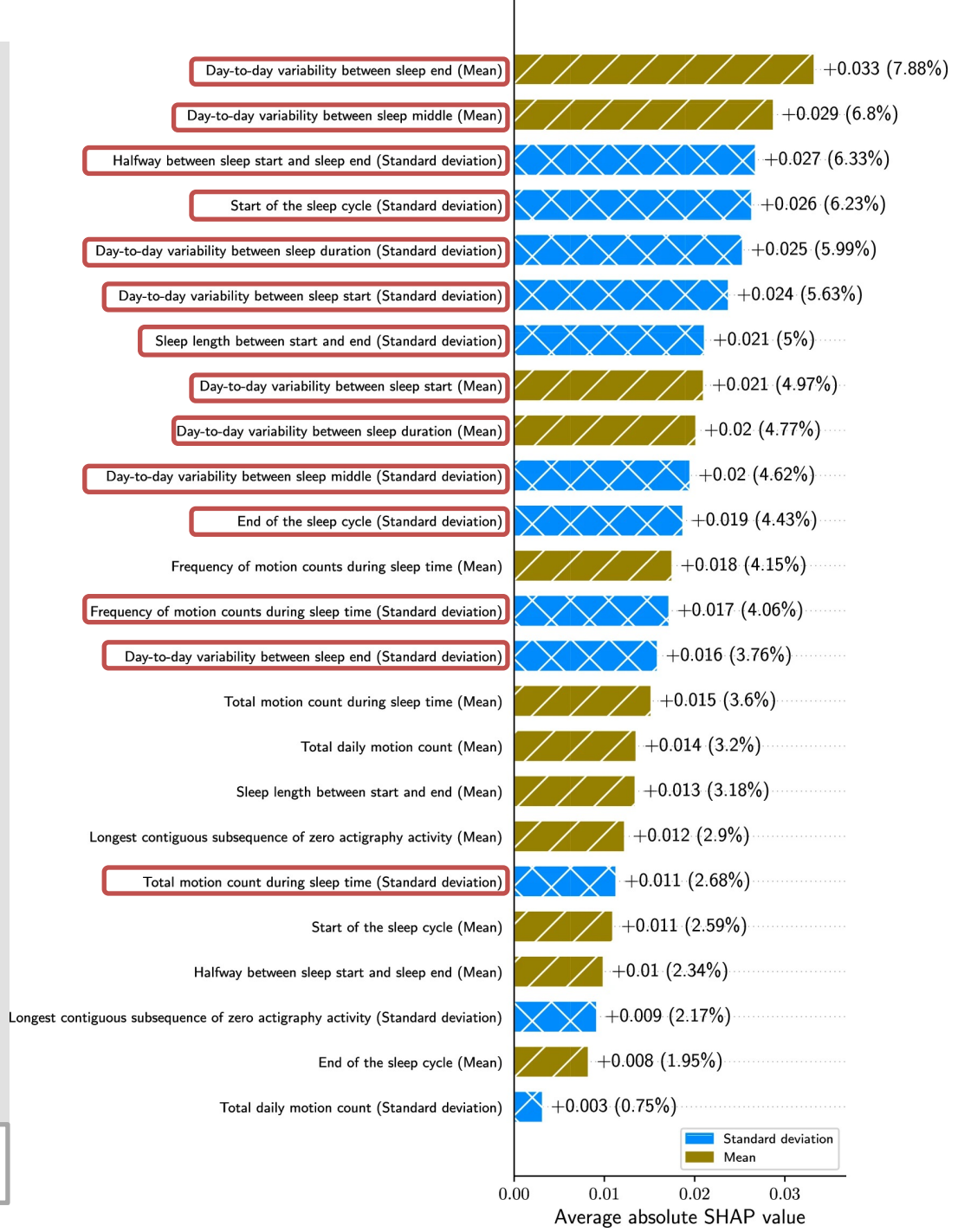
- **Total steps** is the most important wearable feature.
- **Women** and **frequent smokers** are more likely to be diagnosed.

➡ **Match literature**

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.

Predict Preterm Birth

- Preterm birth is estimated to cause 1 million deaths per year.
- 665 patients (297 first-time) from across the BJC HealthCare.
- Wearables and patient characteristics can be used to predict preterm birth.
  - AUROC: 0.722 (0.692–0.753)
- Strongest wearable predictors relate to **variability** in sleep patterns.



Warner BC, Zhao P, Herzog ED, Frolova AI, England SK, Lu C. Validation of sleep-based actigraphy machine learning models for prediction of preterm birth. *npj Women's Health*. 2025;3(1):40.

Three Frontiers in AI for Health

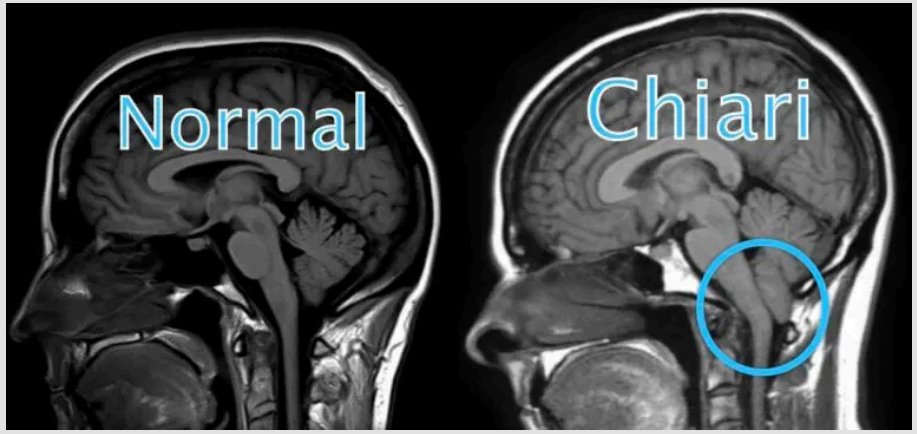
- Predicting clinical outcomes using wearables
- **Phenotyping complex diseases and health service utilization**
- Foundation models for multimodal health data

Chiari Type-1 Malformation (CM1)

- CM1 is a **heterogeneous** disease
  - an abnormality in the back of the head where the brain and spinal cord meet
  - ≥5 mm of cerebellar tonsillar ectopia below the foramen magnum on brain MRI scan

- Congenital disease affecting 3.6% of the population

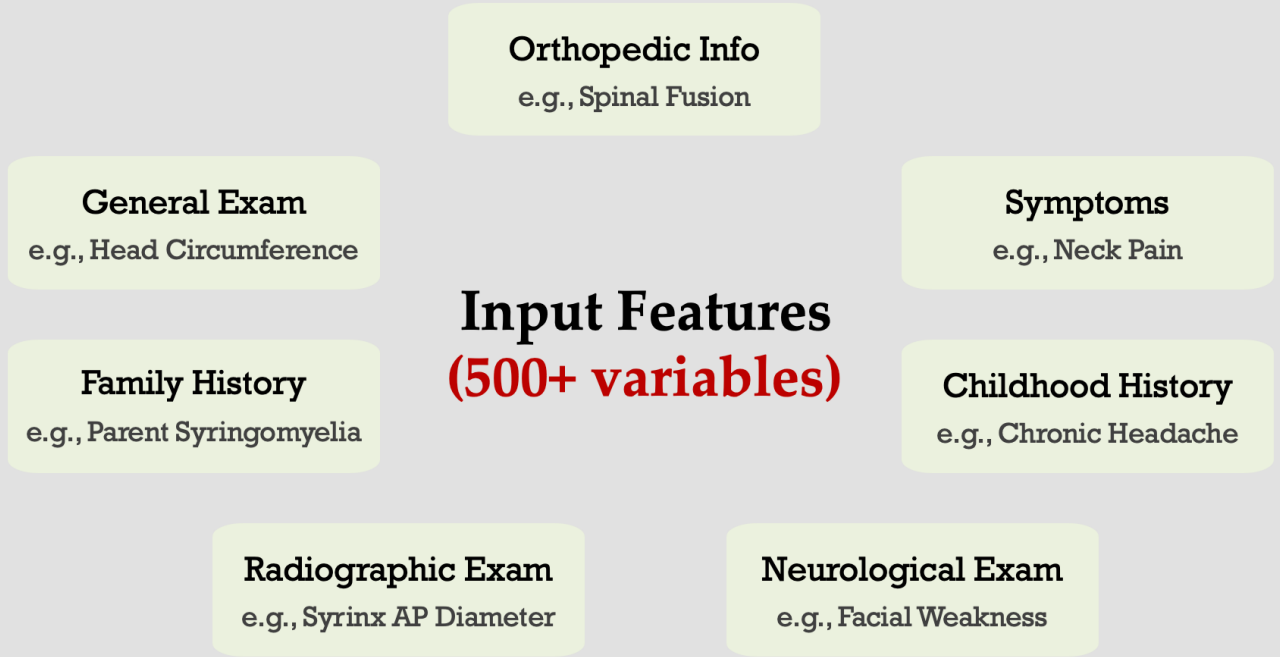
- **Significant differentiation in clinical presentation and natural history**



Phenotype Chiari Malformation Type 1 (CM1)

- CM1 likely reflects multiple clinical phenotypes rather than a single disease entity.
- Goal: Identify **distinct, clinically meaningful** phenotypes

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



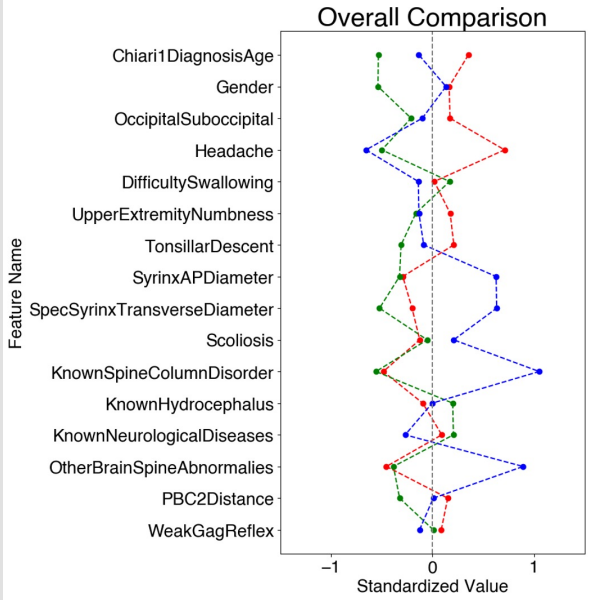
Work with Dave Limbrick, Jacob Greenberg, Sean Gupta (Neurosurgery), Ziqi Xu (CSE)

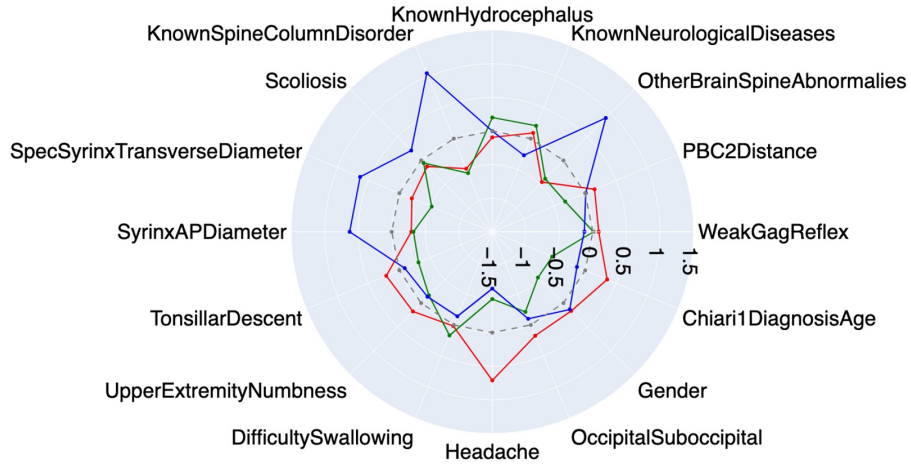
Clustering

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

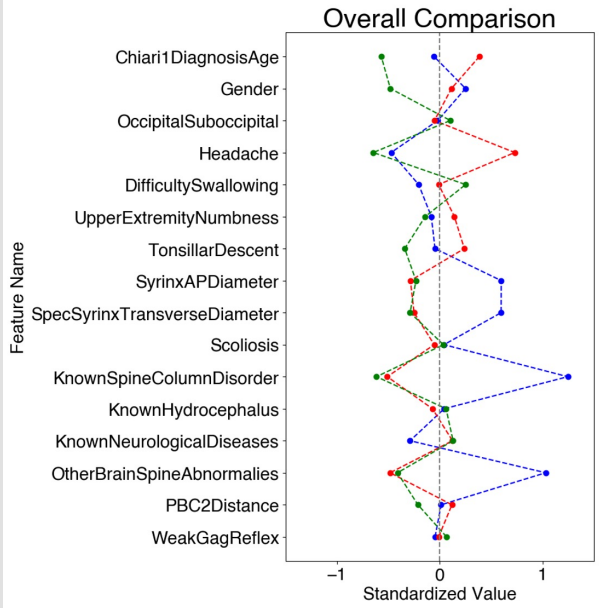
Gupta VP, Xu Z, Greenberg JK, Strahle JM, Haller G, Meehan T, Roberts A, Limbrick DD Jr, Lu C. Using artificial intelligence to identify three presenting phenotypes of Chiari type-1 malformation and syringomyelia. *Neurosurgery*. 2024 Nov 18.

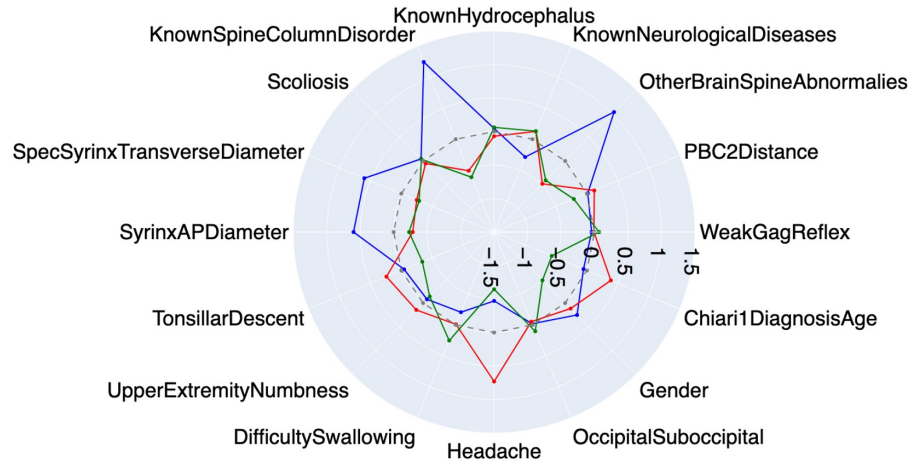
Curve and Radar Plot for Internal Cohort





Curve and Radar Plot for External Cohort





Maternal and Child Health Services Utilization

- Sub-Saharan Africa (SSA) accounts for >80% of the global under-five mortality (U5M) rates.
- **Maternal and Child Health services** are critical in reducing U5M rates.
- Goal: Identify **distinct, actionable** co-utilization patterns of maternal and child health services, and associated socio-economic disparities.

Dataset

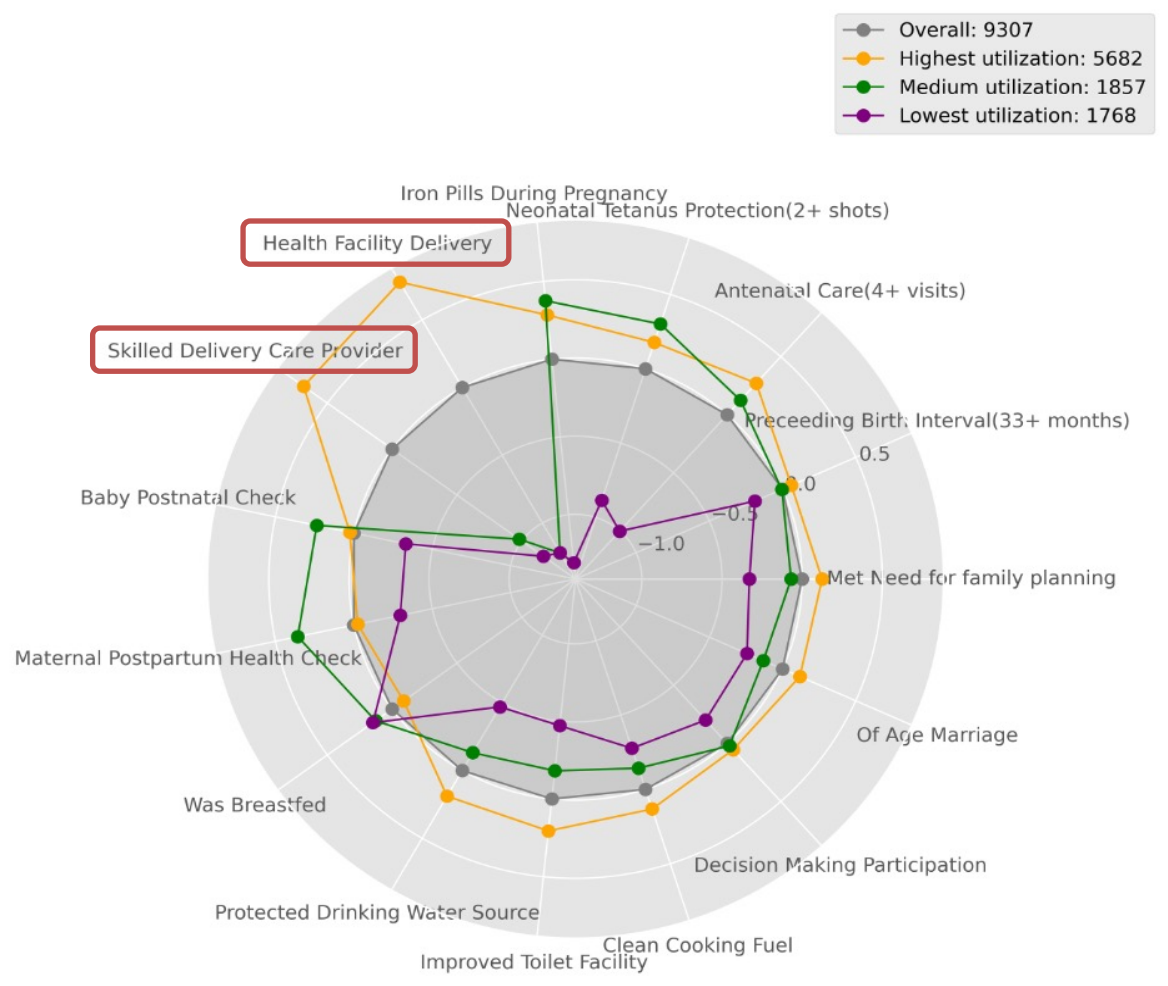
- Nationally representative demographic and health surveys (DHS) from **31 SSA countries** (2014 – 2024)
- **9307** mother-reported most-recent births that ended in U5M

Indicators of maternal and child health services and protective practices

#	Utilization Indicator	% Yes	#	Utilization Indicator	% Yes
1	Met Need for family planning	19.2%	9	Clean fuels for cooking	7.8%
2	Antenatal care (4+ visits)	51.7%	10	Improved sanitation facility	41.7%
3	Iron pills during pregnancy	72.8%	11	Piped water / Protected water source	65.7%
4	Neonatal tetanus protection (2+ shots)	46.0%	12	No tobacco during pregnancy	97.6%
5	Health facility delivery	61.5%	13	Married at mature age	43.0%
6	Skilled provider delivery	61.0%	14	Decision making participation	49.7%
7	Post-partum visits for women	19.5%	15	Child was breastfed	70.1%
8	Post-natal visit for babies	19.9%	16	Child spacing (33+ months)	38.0%

Work with Fred Ssewamala (Social Work), Claire Najjuuko (DCDS)

Service Utilization Patterns



16 utilization indicators of maternal and child health services

Latent Class Analysis

Three distinct utilization groups

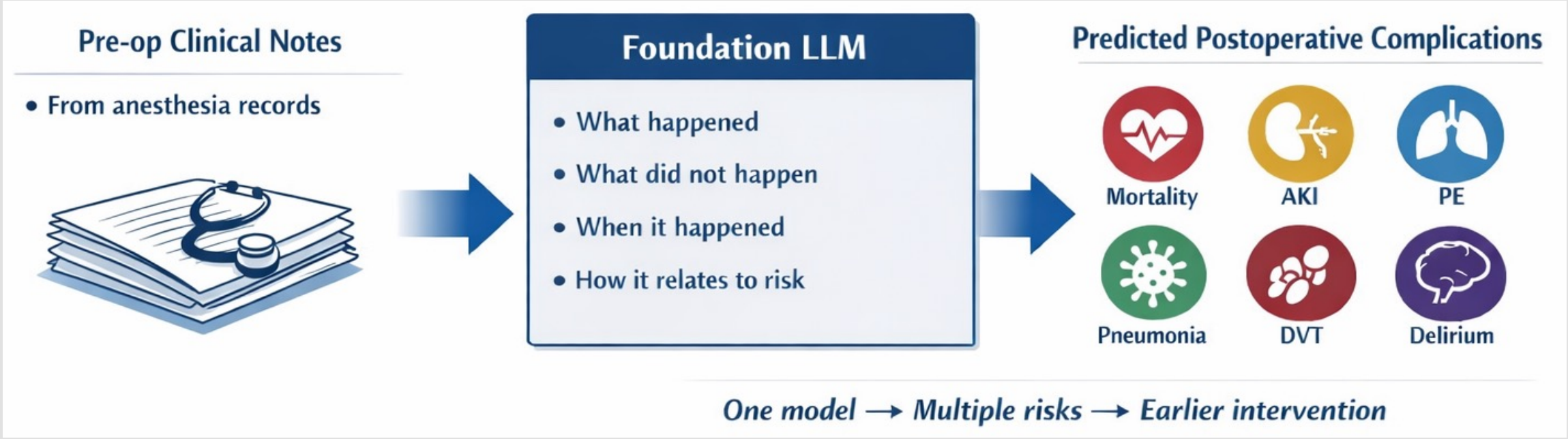
Najjuuko C, Xu Z, Kizito S, Lu C, Ssewamala FM. Patterns of maternal and child health services utilization and associated socioeconomic disparities in sub-Saharan Africa. *Nature Communications*. 2025;16:7840.

Three Frontiers in AI for Health

- Predicting clinical outcomes using wearables
- Phenotyping complex diseases and health service utilization
- **Foundation models for multimodal health data**

Predict Postoperative Risks from Clinical Notes

- Clinical notes capture nuanced clinical judgment that structured data cannot.
- **Large Language Models (LLMs)** translate narrative documentation into risk signals.

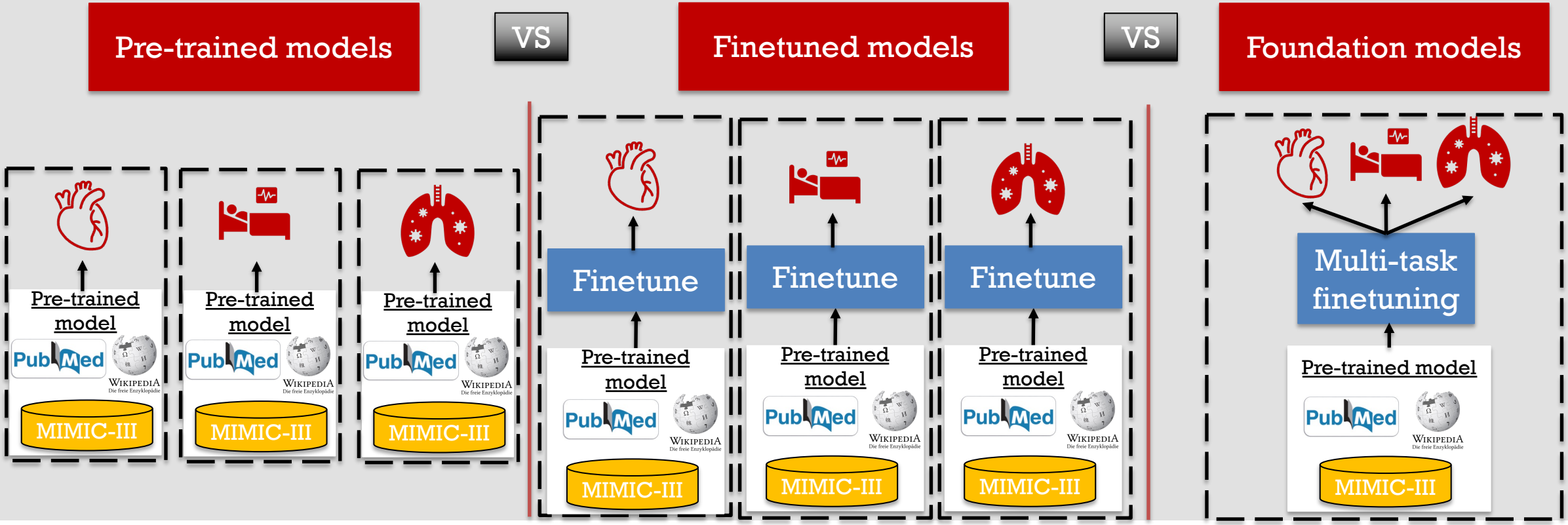


- Dataset: 84,875 pre-operative anesthesia notes from Barnes-Jewish Hospital (BJH)

Work with Joanna Abraham, Thomas Kannampallil (Anesthesiology), Charles Alba (DCDS)

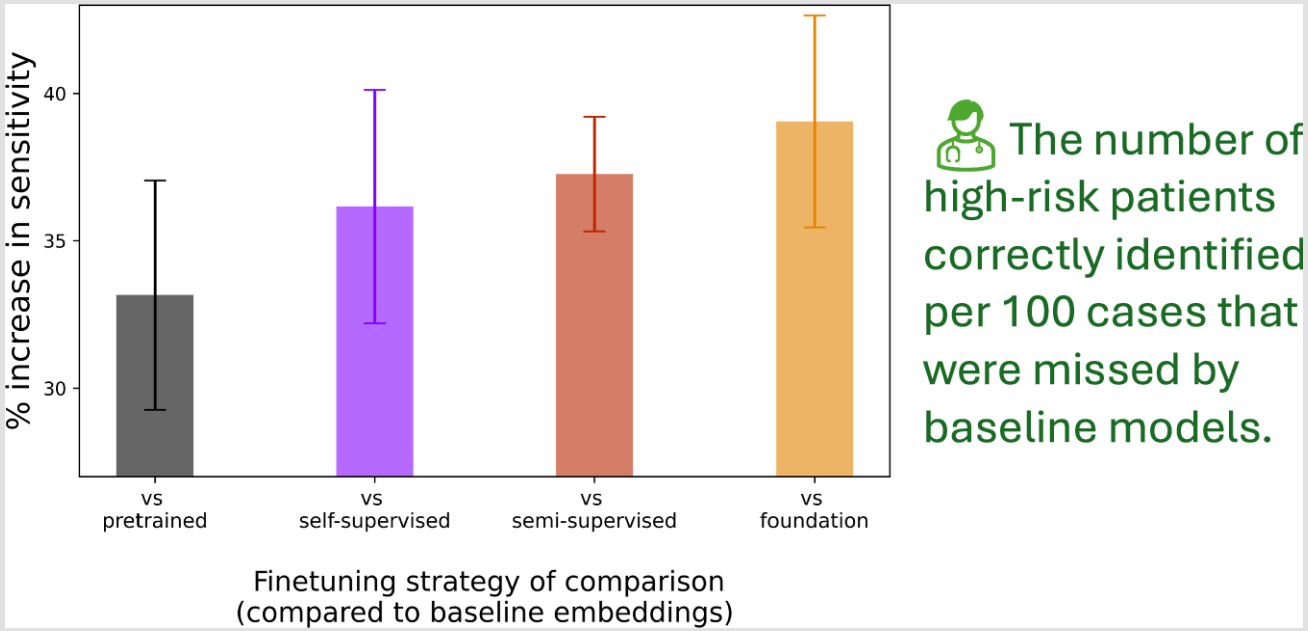
LLMs for Predicting Postoperative Risks

1. Do **pre-trained** LLMs outperform traditional NLP?
2. Does **fine-tuning** to surgical notes improve performance?
3. Can a single **foundation** model predict multiple risks?



Foundation LLMs Improve Surgical Risk Prediction

- Pretrained: **+33** percentage point increase over NLP in sensitivity (at 95% specificity)
- **One** foundation model predicts **multiple** complications
  - **Outperforms** individually fine-tuned models
  - **Simplifies** clinical deployment



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

SurgicalPLAN

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).

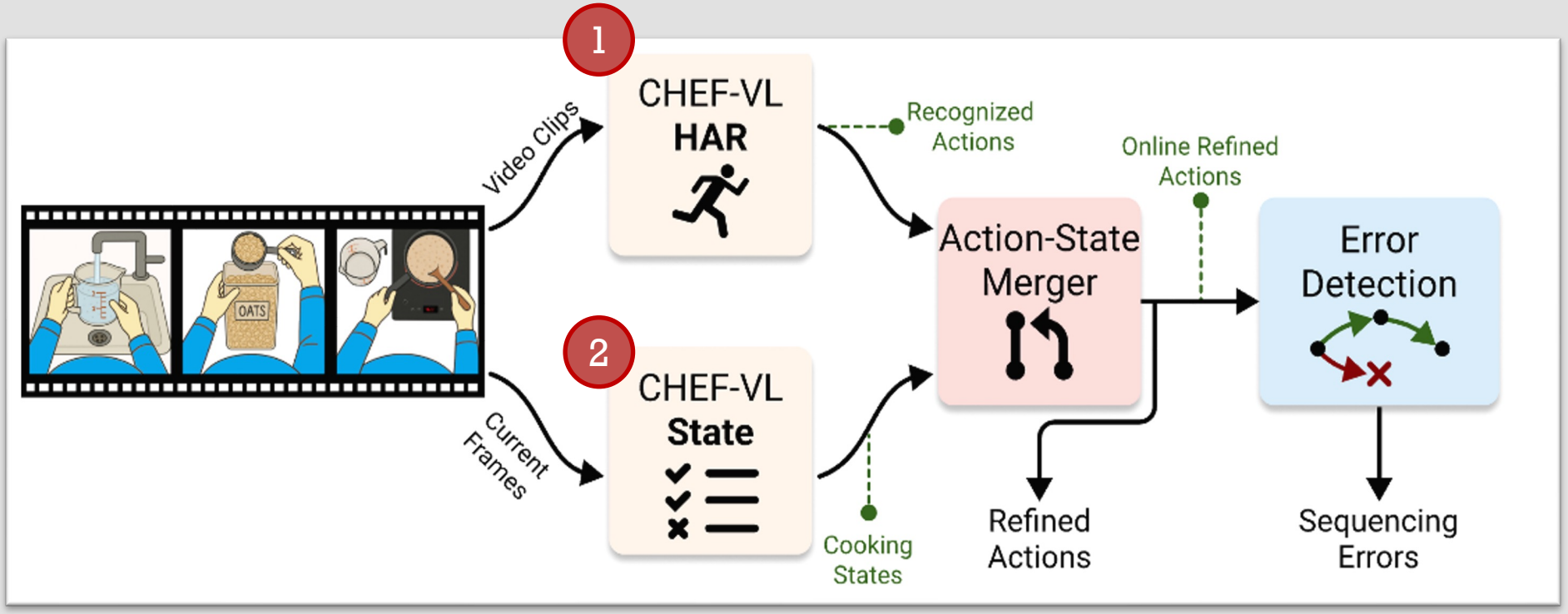
Detecting Cognitive Errors

- Individuals with **Subjective Cognitive Decline (SCD)**
  - Struggle with everyday tasks
  - **1 in 9** people over 45 experience SCD
- **Smart Kitchen**: AI-Enabled Cognitive Support for Everyday Tasks
  - Collected videos of participants cooking oatmeal
  - 107 participants (27: self-reported SCD, 22: mild cognitive impairment).



Work with Lisa Connor, Carolyn Baum (Occupational Therapy), Ruiqi Wang (CSE)

Activity Recognition with Vison Language Models



1. **CHEF-VL-HAR** for action recognition
2. **CHEF-VL-State** for state understanding



Performance: Detecting Sequencing Errors

- **Pretrained** VLM provides a strong zero-shot baseline
- **Fine-tuning** drives large gains in error detection accuracy
- Adding **state** information delivers the highest overall performance

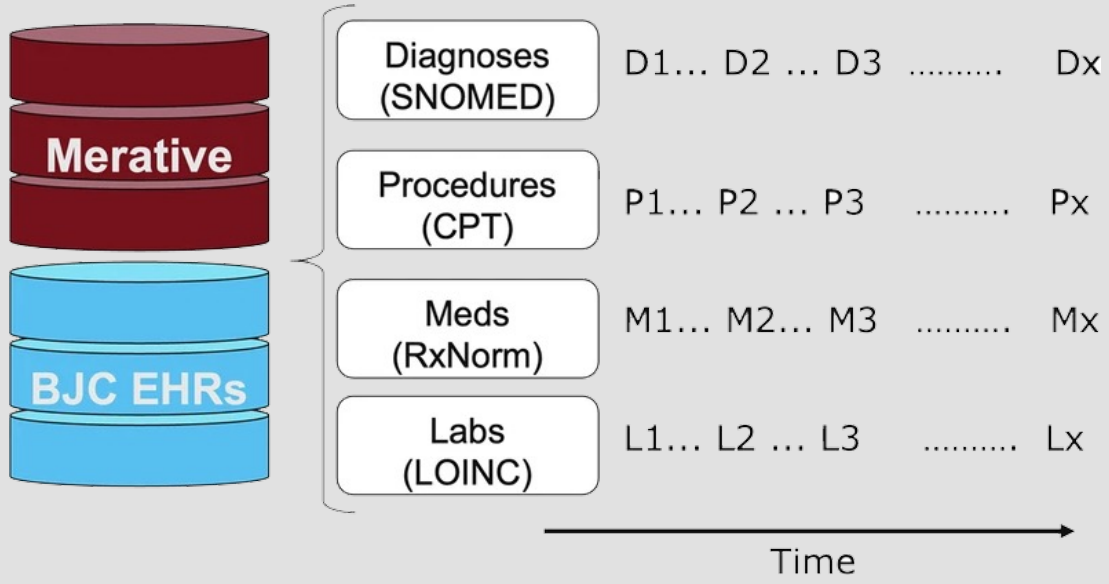
Method	Precision (%)	Recall (%)	F1-Score (%)
LSTR	20.42	60.67	19.07
Qwen2.5-VL	33.00	35.00	17.00
+ Fine-tune → Qwen2.5-VL-Finetune	<u>49.83</u>	<u>70.67</u>	<u>44.50</u>
+ State → **CHEF-VL (Ours)**	**53.00**	**71.67**	**48.43**

Wang R, Gao P, Lynch P, Liu T, Lee Y, Baum CM, Connor LT, Lu C. CHEF-VL: Detecting cognitive sequencing errors in cooking with vision-language models. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2025;9(4):220. doi:10.1145/3770714.

Predict Cervical Spondylotic Myelopathy (CSM)

- CSM is the **leading cause** of spinal cord dysfunction in older adults
- Diagnosis is frequently delayed by **2–6 years** after symptom onset
- Goal: Predict CSM using **longitudinal EHR data** before clinical diagnosis

Data Sources
- Merative MarketScan: ~1.4M patients' longitudinal claims data
- BJC Health: ~500K patients' institutional EHRs



Yakdan S, Warner BC, Ghogawala Z, Ray WZ, Bydon M, Steinmetz MP, Griffey RT, Foraker R, Wilcox A, Lu C, Greenberg JK. Clinically-guided models or foundation models? Predicting cervical spondylotic myelopathy from electronic health records. *npj Digital Medicine*. 2026;9:153.

AI: Transforming Health

- Analyze **multimodal, high-dimensional** data
- **Predict** outcomes → enable personalized care
- **Interpret** models → deliver actionable insights
- **Phenotype** diseases → support precision interventions



AI for Health: Looking Ahead

- **Human–AI collaboration**
  - Clear explanations, calibrated uncertainty
  - Real workflow value
- **Model robustness**
  - Generalize across sites
  - Stay reliable over time
- **Stronger partnerships**
  - AI + clinicians + public health experts
  - Co-design solutions that matter

AI for Health Institute

aihealth.wustl.edu

CSE 4109: Interdisciplinary Class on AIHealth



- **Projects 60%**
  - Proposal and presentation: 10%
  - Demo I: 5%
  - Demo II: 5%
  - Final report and demo: 40%
- **Critiques 35%**
- **Participation 5%**

- **Interdisciplinary**: co-taught by experts across AI and health
- **Research** experience on AIHealth

Critiques

- Half-page **critiques** of research papers
- Due by 10am on class day
- Back-of-envelop - NOT whole essays
- See requirements on class website

Project

- **Two** students per team
  - Need TA permission for a team of a different size.

- Propose a AIHealth project
- Develop AI models
- Train and validate models using clinical data
- Write a paper
- Demos

- Opportunity to contribute to **AIHealth** research

Logistics



- Guidelines and slides are on the class **homepage**.
- All work will be submitted through **Canvas**.
- Find critique and demo and due dates on the class **schedule**.

http://www.cse.wustl.edu/~lu/cse4109/

Support



- Prof. Chenyang Lu lu@wustl.edu
- Peiqi Gao peiqi@wustl.edu
- Claire Najjuuko c.najjuuko@wustl.edu

- Email for appointments

Next Class



- Project guidelines and topics