AI for Health

Translating Wearable Data into Clinical Impact

Chenyang Lu

Director, AI for Health Institute

Fullgraf Professor

Computer Science · Anesthesiology · Medicine · Neurosurgery · Psychiatry · Public Health

Washington University in St. Louis

Wearables Capture Health Beyond the Clinic

Wearables are already mainstream

- **Commonly available**
  - Step count
  - Heart rate
  - Sleep stages
- **Expanding form factors**
  - Watches, bands, and rings
  - Glasses, earbuds, and smart clothing

Sensing is becoming richer

- **More sensing modalities**
  - Oxygen saturation (SpO2)
  - Skin temperature and breathing rate
  - Heart rate variability and ECG
  - Stress

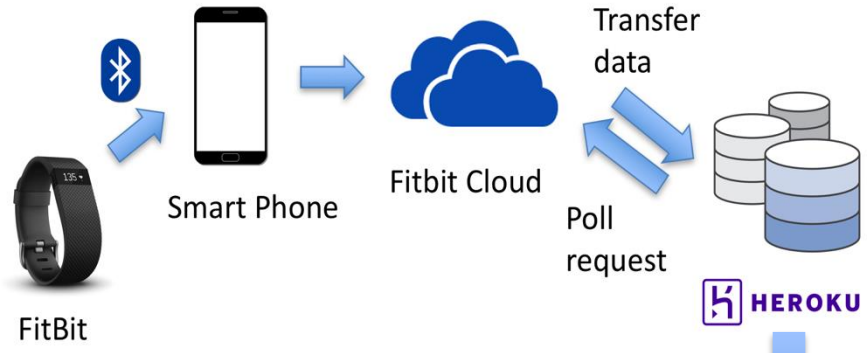
600+ million wearable devices shipped worldwide in 2025

Unobtrusive, longitudinal monitoring outside hospitals

Source: IDC Worldwide Quarterly Wearable Device Tracker, 2025.

Wearable AI for Surgical Decision Support

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





Machine learning

Predictions

Clinical decisions and intervention

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

Robust Feature Engineering Pipeline



Daily Feature Extraction



Step Count

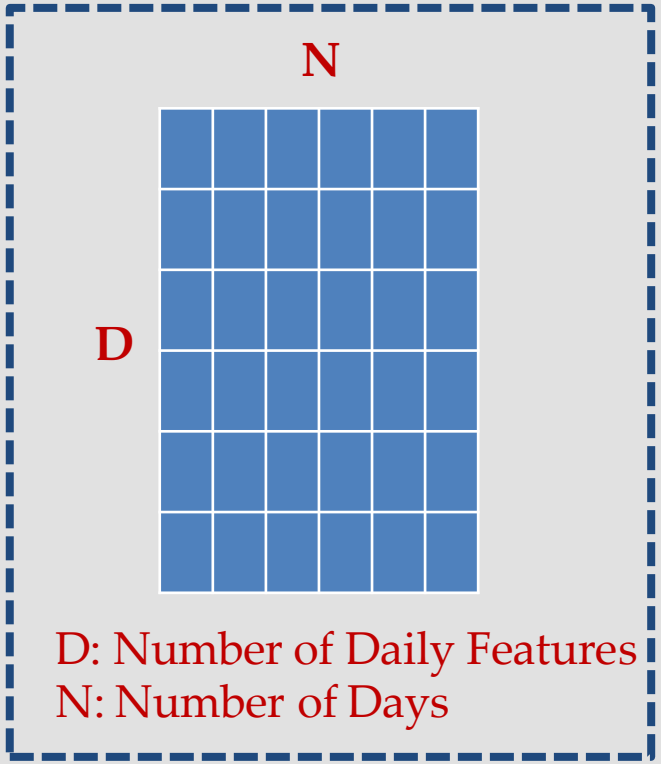


Heart Rate

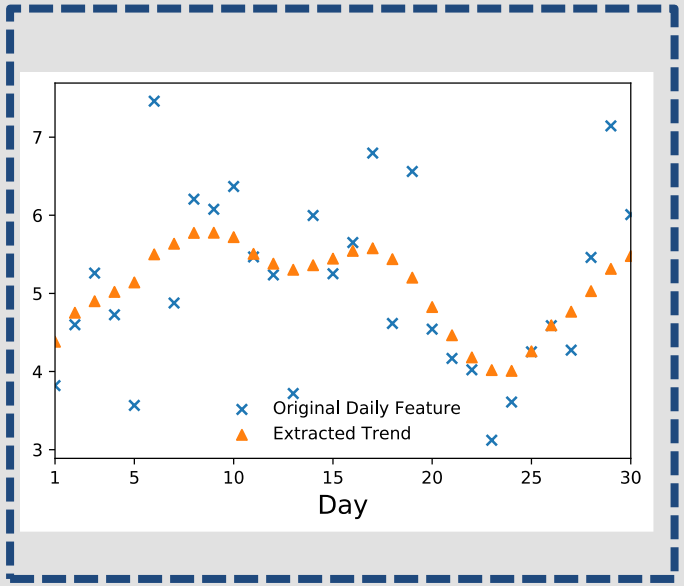


Sleep Stage

Wearable Time Series



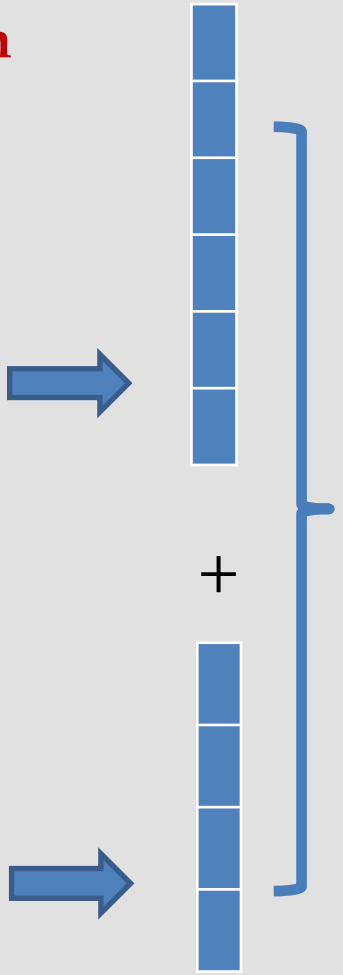
High-level Feature Extraction







Clinical Characteristics



Wearables Improve Surgical Risk Prediction

Machine learning outperforms standard surgical risk scores.

x2 AUPRC

x3 sensitivity at the same specificity

Data Source	AUROC	AUPRC	Sensitivity	Specificity
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

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. **Distinguished Paper Award**

Roadmap to Clinical Impact

Promising pilot studies → TRANSLATION → **Scalable clinical impact**

Complexity

- Meet clinical standards
- Predict multidimensional outcomes

Scale

- Evidence based on large sample sizes

AI-Clinician Collaboration

- Calibrate uncertainty
- Support clinical decisions

Personalized Treatment

- Predict individual response
- Guide treatment choices

Roadmap to Clinical Impact

Promising pilot studies → TRANSLATION → **Scalable clinical impact**

Complexity

- Meet clinical standards
- Predict multidimensional outcomes

Scale

- Evidence based on large sample sizes

AI-Clinician Collaboration

- Calibrate uncertainty
- Support clinical decisions

Personalized Treatment

- Predict individual response
- Guide treatment choices

Toward Standard of Care

Patient and procedure variability

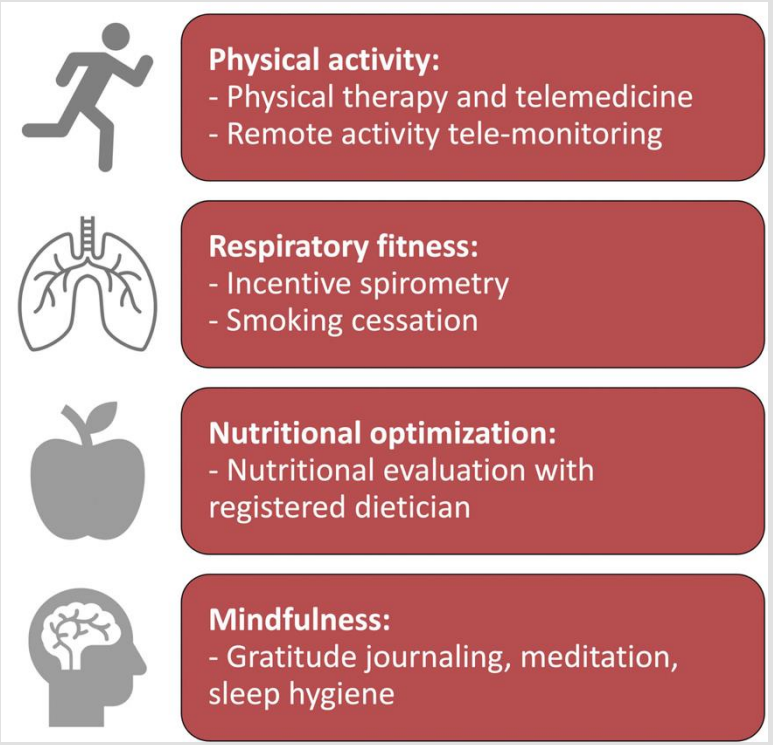
- Generalize across procedures and patient subgroups
- Build robust models for heterogeneous populations

From trials to standard care

- Sustain wearable use in real-world settings
- Handle missing and noisy sensor data

From prediction to intervention

- Turn predictions into actionable clinical insight
- Personalize treatment to each patient



Surgical Prehabilitation and Readiness (SPAR) Program at WashU Surgery

Clinical Outcomes Are Multidimensional

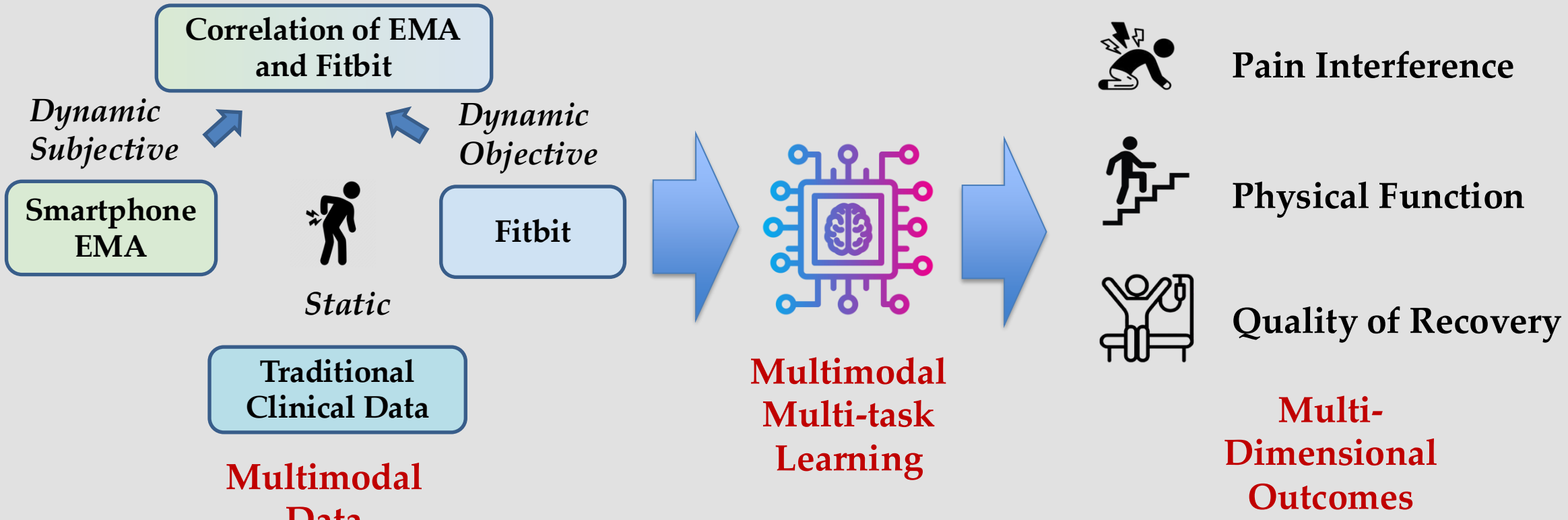
- Low-back pain affects 1.36 million people and costs $135 billion in the U.S.
- Lumbar spine surgery can help, yet recovery varies widely
- Interrelated outcomes: pain, physical function, mental health



Goal: Predict **multi-dimensional** recovery outcomes

Work with Jacob Greenberg (Neurosurgery), Ziqi Xu (CSE)

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).

Roadmap to Clinical Impact

Promising pilot studies → TRANSLATION → **Scalable clinical impact**

Complexity

- Meet clinical standards
- Predict multidimensional outcomes

Scale

- Evidence based on large sample sizes

AI-Clinician Collaboration

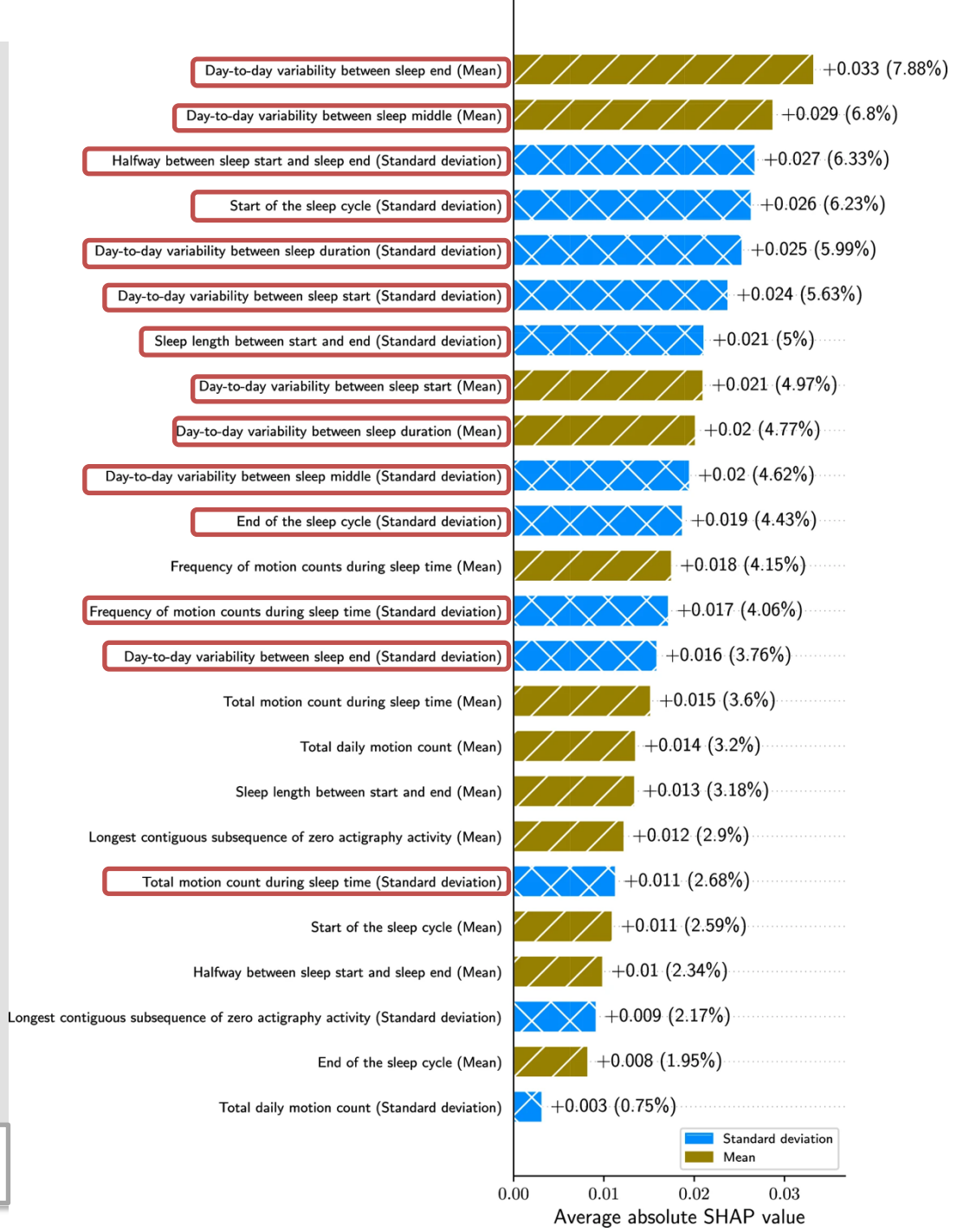
- Calibrate uncertainty
- Support clinical decisions

Personalized Treatment

- Predict individual response
- Guide treatment choices

Wearables Predict Preterm Birth

- Preterm birth causes an estimated 1 million deaths each year.
- Study: 665 BJC HealthCare patients, including 297 first-time mothers.
- Wearables plus patient characteristics predicted preterm birth (AUROC 0.722, 95% CI 0.692–0.753).
- Variability in sleep patterns ranked among the strongest wearable predictors.



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.

Scale Changes What Wearable AI Can Learn

- Depression affects 3.8% of the global population (280 million people; WHO).
- More than half of patients are not recognized or treated.
- Clinical screening is time intensive and costly.
- WearNet detects depression and anxiety from daily wearable signals.
- All of Us cohort: 8,996 participants, including 1,247 positive cases.



Daily wearable signals

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

WearNet: Outcomes and Inputs

What did we detect?

Clinician-recorded diagnoses across 20 depression and anxiety categories

Positive label: at least one of the diagnoses

What data did we use?

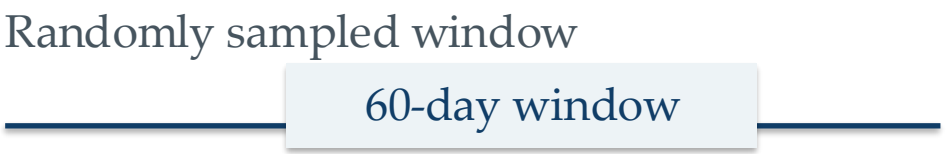
60 days of wearable data
10 daily features of steps and heart rate

Input Data

Participants with a diagnosis



Participants without a diagnosis

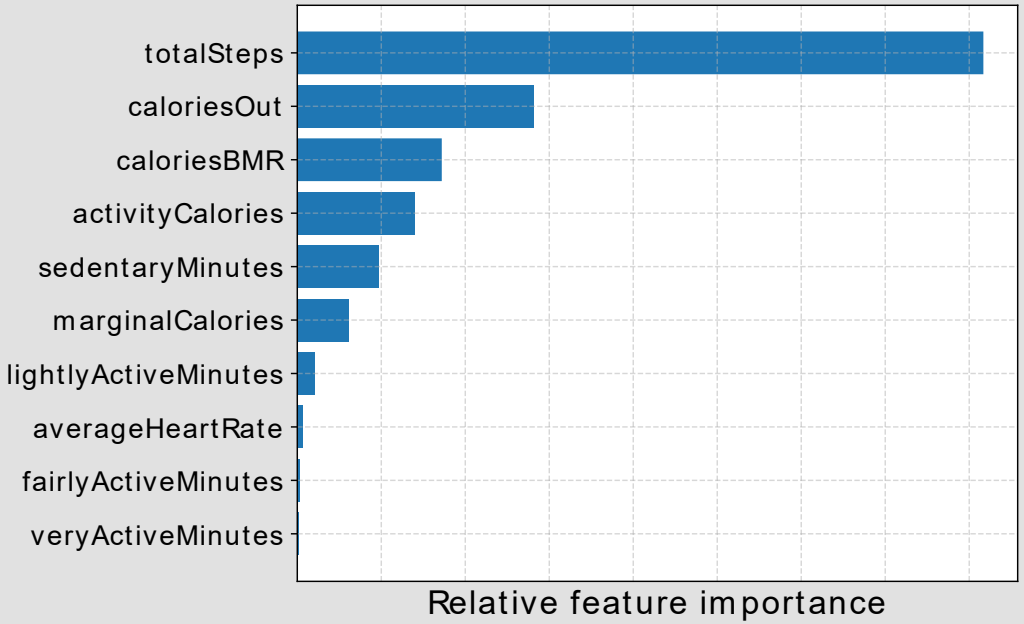


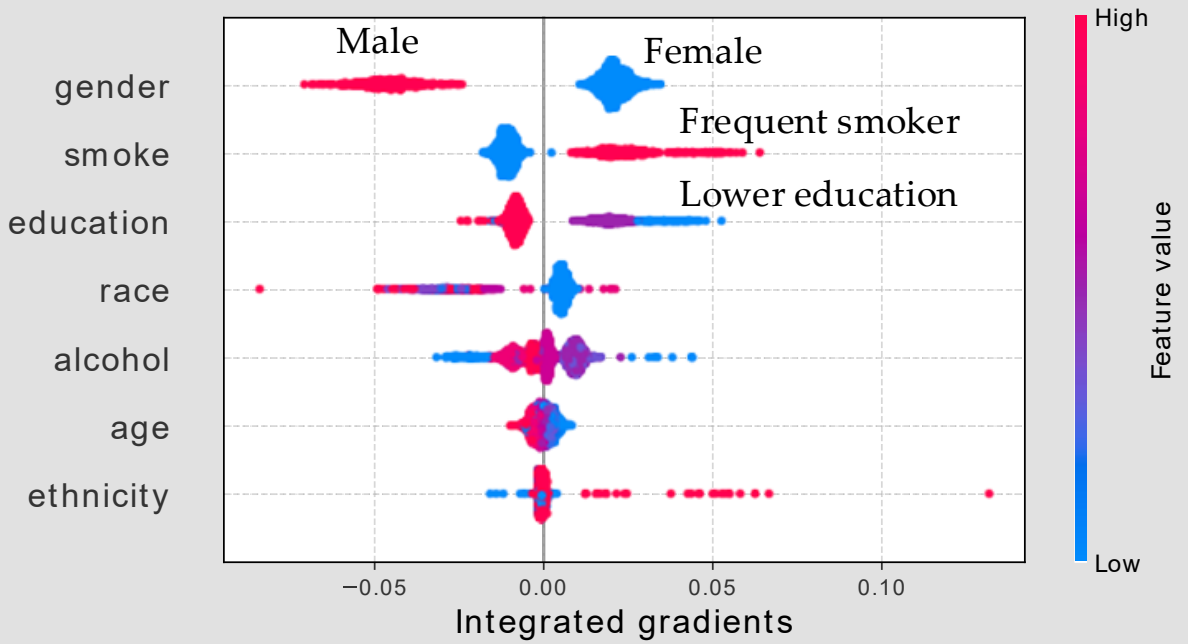
AUROC: 0.717

Dai R, Kannampallil T, Kim S, Thornton V, Bierut L, Lu C. Detecting mental disorders with wearables: a large cohort study. IoTDI, 2023.

Interpreting WearNet Predictions

Feature importance identifies the variables driving the predictions.





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

→ Consistent with literature

Dai R, Kannampallil T, Kim S, Thornton V, Bierut L, Lu C. Detecting mental disorders with wearables: a large cohort study. ACM/IEEE IoTDI. 2023.

Roadmap to Clinical Impact

Promising pilot studies → TRANSLATION → **Scalable clinical impact**

Complexity

- Meet clinical standards
- Predict multidimensional outcomes

Scale

- Evidence based on large sample sizes

AI-Clinician Collaboration

- Calibrate uncertainty
- Support clinical decisions

Personalized Treatment

- Predict individual response
- Guide treatment choices

Clinical AI Must Represent Uncertainty



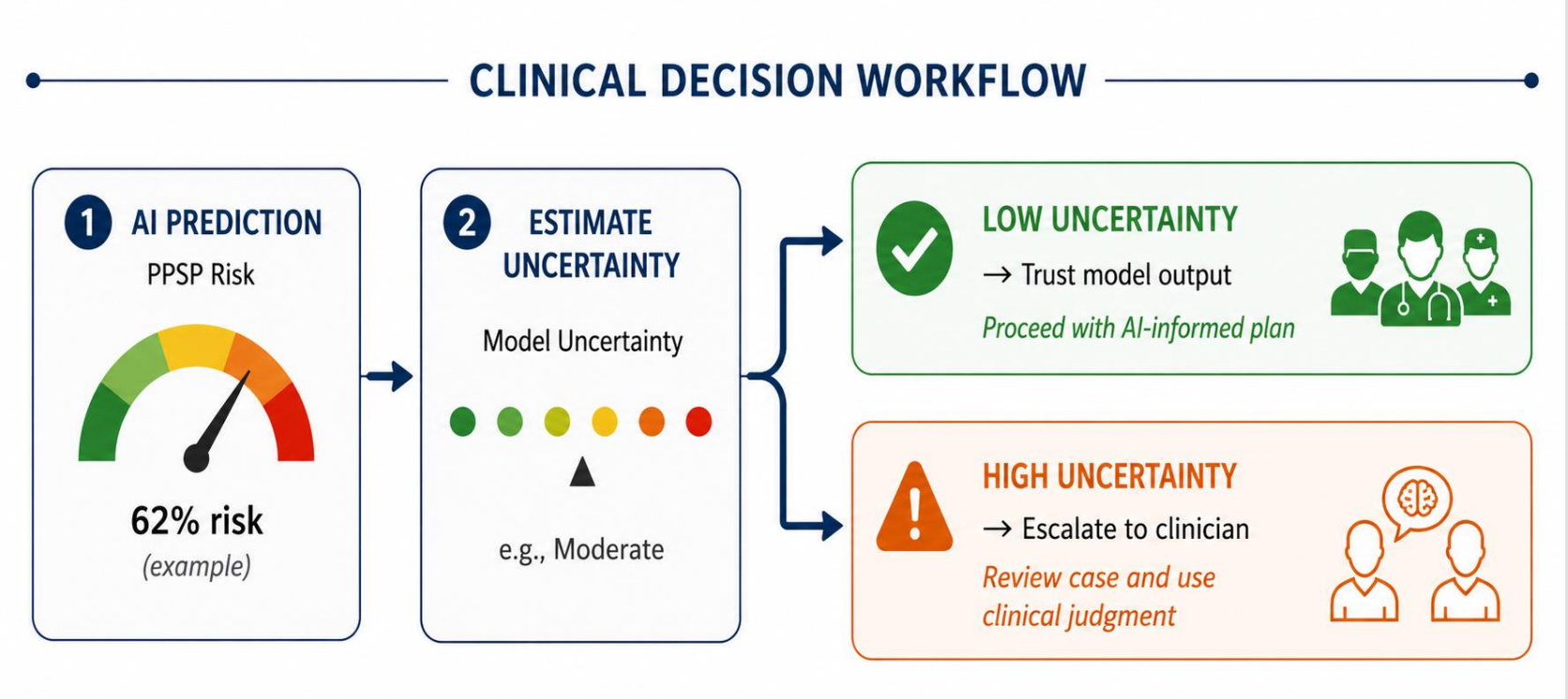
- Persistent pain: pain lasts more than three months after surgery.
- Persistent pain affects 10–35% of patients and about 45 million people each year.
- Static clinical data miss changing factors such as mental health and activity.
- Calibrated uncertainty estimates supports clinician-AI collaboration.

Goal: Uncertainty-aware AI for PPSP prediction

Joint work with Simon Haroutounian (Anesthesiology), Ziqi Xu (CSE)

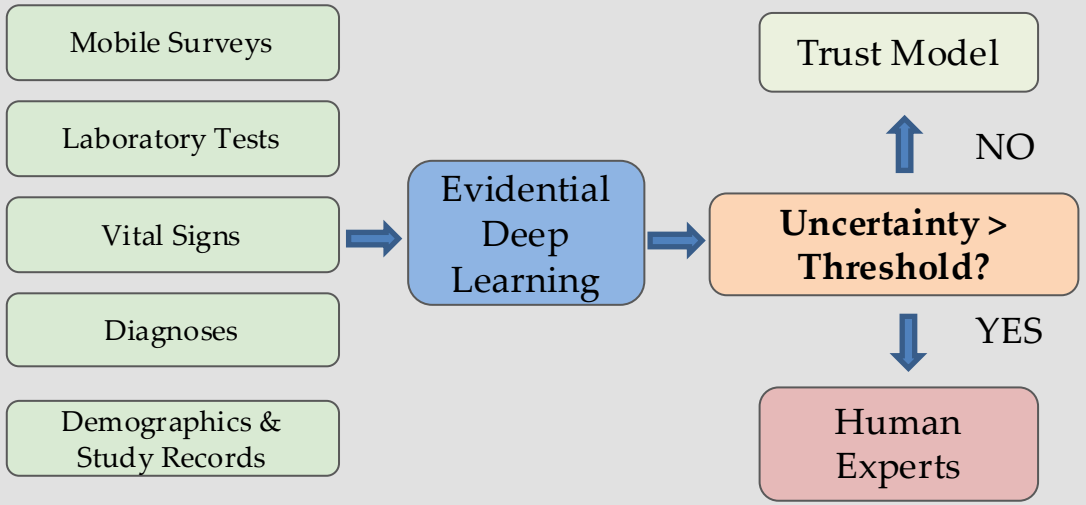
Uncertainty Supports Safer Clinical Decisions

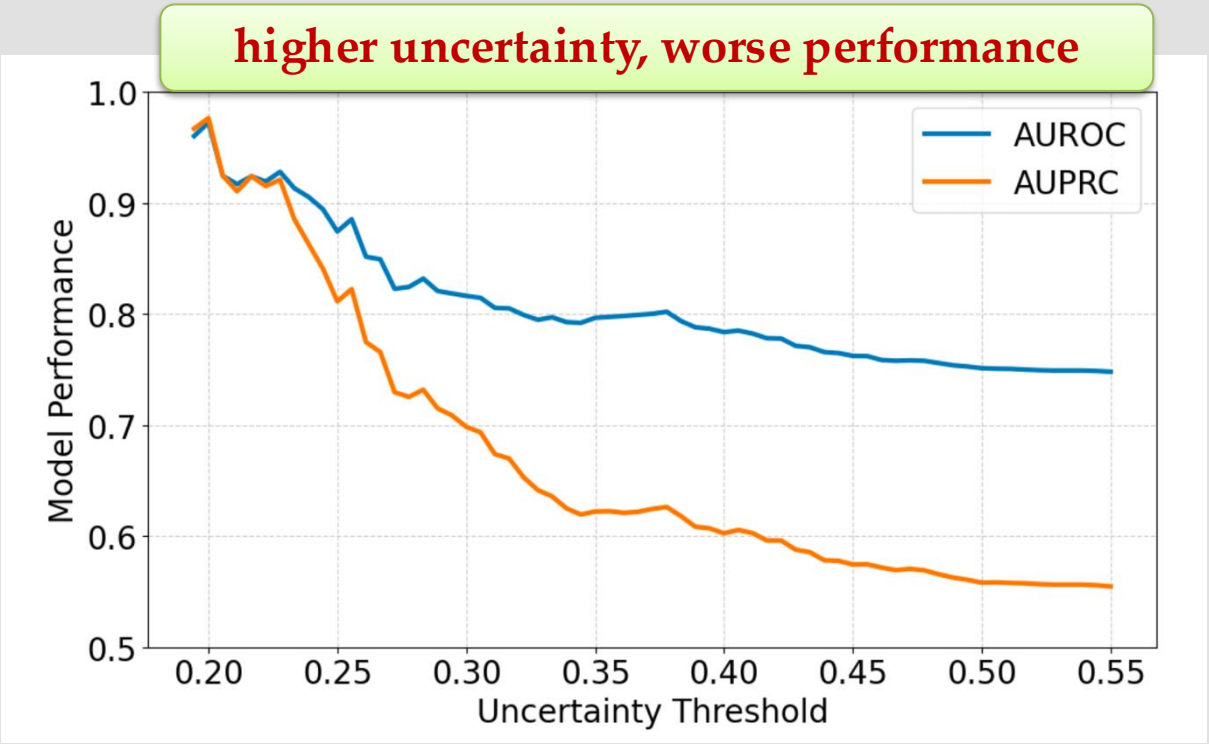
- Complications, recovery, and long-term outcomes involve high-stakes decisions.
- Confidently wrong predictions can mislead clinical decisions.
- Uncertainty helps clinicians decide when to trust the model and when to defer.



Uncertainty Reveals Less Reliable Predictions

- Inputs combine preoperative mobile sensing and clinical data.
- Estimated uncertainty guides when clinicians should rely on the model.





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.

Roadmap to Clinical Impact

Promising pilot studies → TRANSLATION → **Scalable clinical impact**

Complexity

- Meet clinical standards
- Predict multidimensional outcomes

Scale

- Evidence based on large sample sizes

AI-Clinician Collaboration

- Calibrate uncertainty
- Support clinical decisions

Personalized Treatment

- Predict individual response
- Guide treatment choices

From Average Treatment Effects to Individual Prognosis

A randomized trial asks whether a treatment helps, on average.

Population-level evidence

Does depression improve more with the intervention than with usual care?

Individual-level prediction

What depression outcome is likely for this patient with the intervention or usual care?

The opportunity: use trial data to predict individual outcomes.

Work with Thomas Kannampallil (Anesthesiology), Jun Ma (UIC), Ruixuan Dai (AIHealth)

Case Study: Predicting Depression Remission



Trial: 106 adults with depression and obesity (NCT03841682)

Trial intervention

Integrated collaborative care for depression and obesity, including behavioral activation over 6 months.

71 intervention; 35 usual care

Predicted outcome

Depression remission at 6 months (SCL-20 < 0.5)

Predict remission within each trial group

Goal: predict remission using baseline data and the first 2 months of Fitbit data.

Dai R, Kannampallil T, Zhang J, Lv N, Ma J, Lu C. Multi-Task Learning for Randomized Controlled Trials: Predicting Depression with Wearable Data. IMWUT. 2022;6(2):50.

Clinical History and Daily Life Inform Prediction

Predict six-month remission using baseline and early wearable data.

Baseline

Clinical information

Age, anxiety, and other baseline characteristics

First 2 months

Wearable data

Heart rate, sleep, steps

At 6 months

Depression remission

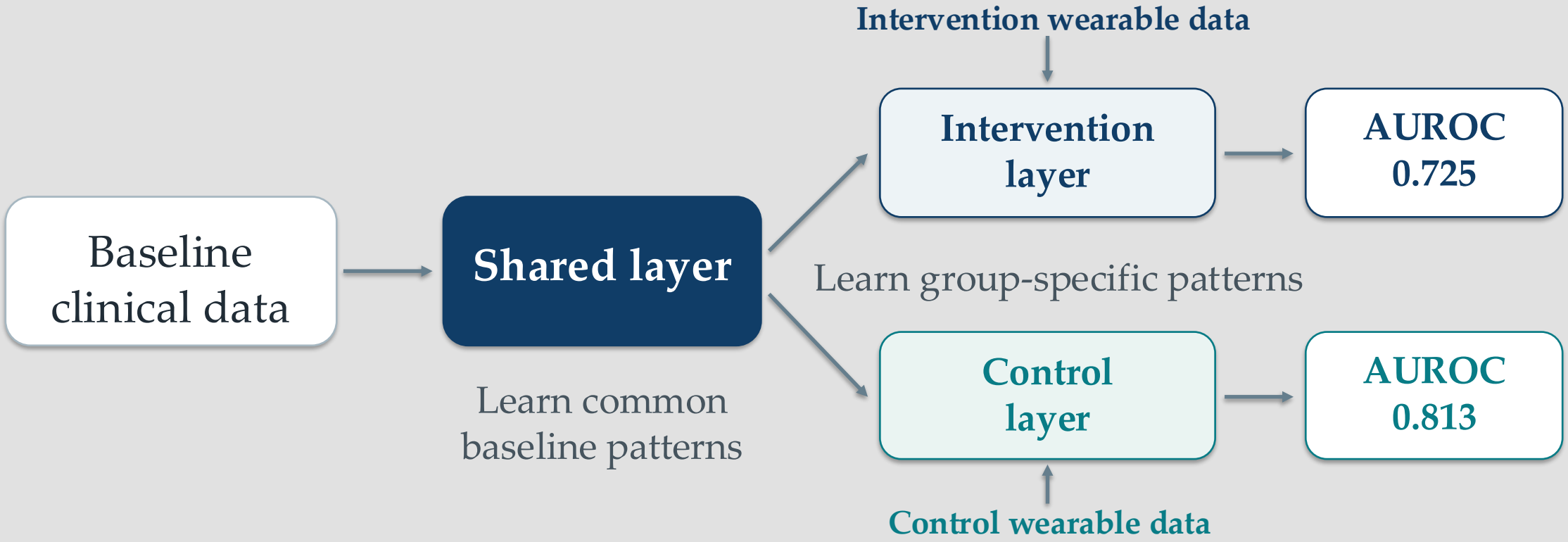
SCL-20 < 0.5
Primary endpoint

Clinical goal: inform follow-up using each patient's early data.

Dai R, Kannampallil T, Zhang J, Lv N, Ma J, Lu C. Multi-Task Learning for Randomized Controlled Trials: Predicting Depression with Wearable Data. IMWUT. 2022;6(2):50.

One Model for Both Trial Groups

Predict six-month depression remission within each trial group



Dai R, Kannampallil T, Zhang J, Lv N, Ma J, Lu C. Multi-Task Learning for Randomized Controlled Trials: Predicting Depression with Wearable Data. IMWUT. 2022;6(2):50.

Predict Individual Response to Cognitive Behavioral Therapy

2,295
students

26
U.S. colleges

2 years
follow-up

Intervention and population

Digital CBT guided self-help for undergraduates at risk for, or meeting criteria for, anxiety, depression, or eating disorders.

Prediction before treatment

Who will be free of all target disorders at two years?

Prevention or remission, based on baseline information.

Work with Denise Wilfley (Psychiatry), Ellen Fitzsimmons-Craft (Psychology), Michelle Newman, Adam Calderon (PSU), Daoyi Zhu (AIHealth)

Fewer Inputs, Lower Assessment Burden

Model development: 1,472 students across 13 colleges

863

baseline questionnaire items

→

37

predictors retained

Four Predictor Domains	
Demographics	Mental health symptoms
Treatment targets	Treatment motivation

Fewer questions could make the tool easier to use across campuses.

Calderon A, Zhu D, Zainal NH, Lu C, Fitzsimmons-Craft EE, Wilfley DE, Eisenberg D, Taylor CB, Newman MG. Psychotherapy and Psychosomatics. 2026. doi:10.1159/pps/aemag007.

External Validation in 13 Other Colleges

DEVELOPMENT

1,472 students

13 colleges

EXTERNAL VALIDATION

823 students

13 separate colleges

0.758

External-validation AUROC

Does the model predict outcomes at colleges that were not used for model development?

Generalizability across colleges

Calderon A, Zhu D, Zainal NH, Lu C, Fitzsimmons-Craft EE, Wilfley DE, Eisenberg D, Taylor CB, Newman MG. Psychotherapy and Psychosomatics. 2026. doi:10.1159/pps/aemag007.

Wearable AI Can Transform Health

Continuous, real-world data create opportunities across clinical domains.

1

SURGERY

Predict complications and recovery

2

MENTAL HEALTH

Detect risk and personalize treatment response

3

WOMEN'S HEALTH

Identify risk earlier outside the clinic

Earlier insight **Better prognosis** **Treatment tailored to the patient**

A Roadmap from Wearable Data to Clinical Impact



1

Evidence at scale

Large, diverse cohorts
External validation

2

Robust deployment

Generalize across sites
Stay reliable over time

3

Clinician–AI collaboration

Calibrated uncertainty
Fit clinical workflows

4

Personalized care

Predict individual outcomes
Guide treatment decisions

Translation is the next breakthrough.

Let’s Build the Translation Path Together

Combining clinical and AI expertise to accelerate translation

AI for Health Institute

aihealth.wustl.edu