CPSL
Cyber-Physical Systems Laboratory

Internet of Medical Things

Precision Medicine with AI and Wearables

Chenyang Lu
Fullgraf Professor of Computer Science & Engineering
Professor of Anesthesiology and Medicine
AIM Institute (AI and IoT for Medicine)

Washington University in St. Louis

AIM: AI and IoT for Medicine

- AI: Data-driven tools for medicine
  - Phenotype complex diseases
  - Predict clinical outcomes
  - Discover risk factors
  - Support clinical decisions

- Extract knowledge from diverse data

Electronic Health Record (EHR)
- **Collected in hospitals**
- **Complex and high-dimension data**

Internet of Things (IoT)
- **Longitudinal monitoring outside hospitals**
- **Noisy and lossy 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

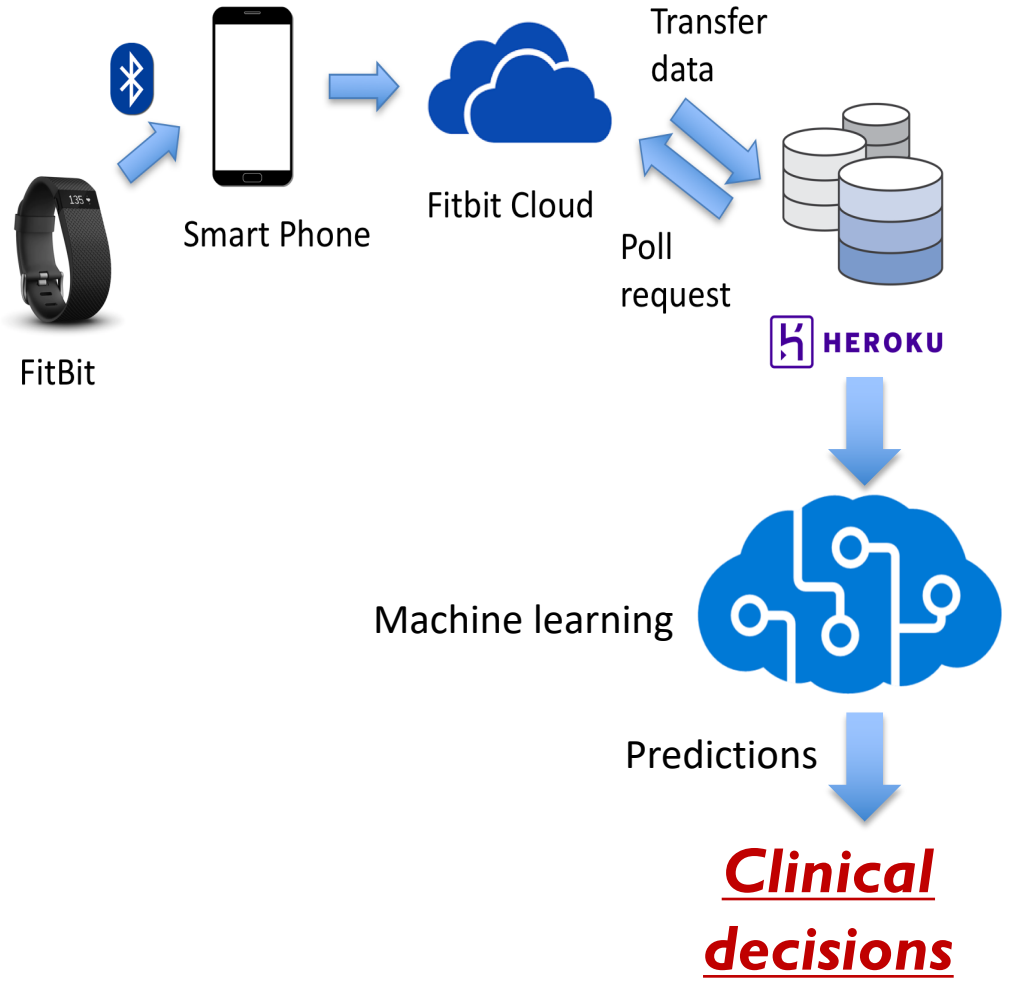




Unprecedented monitoring capability outside hospitals!

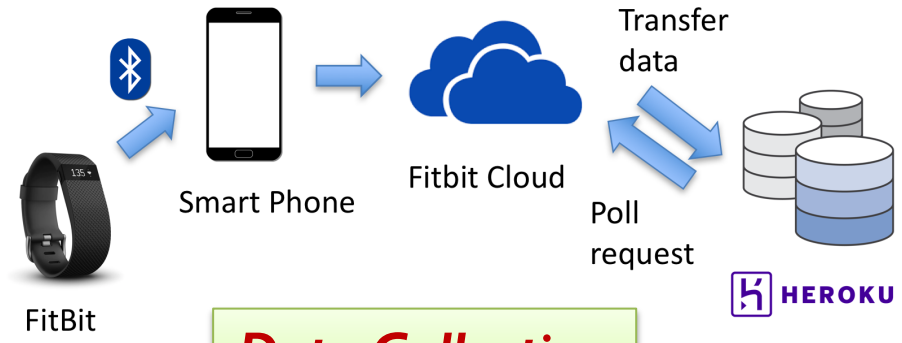
IoMT: Internet of Medical Things

- **Wearables**: wristband, watch, ring…
  - Long-term, non-obtrusive monitoring
- **Connectivity**: Bluetooth, WiFi, cellular
  - Real-time monitoring and intervention
- **Cloud**: computing and storage
  - Scalable to large population
- **Analytics**: machine learning
  - Predict outcomes and support intervention

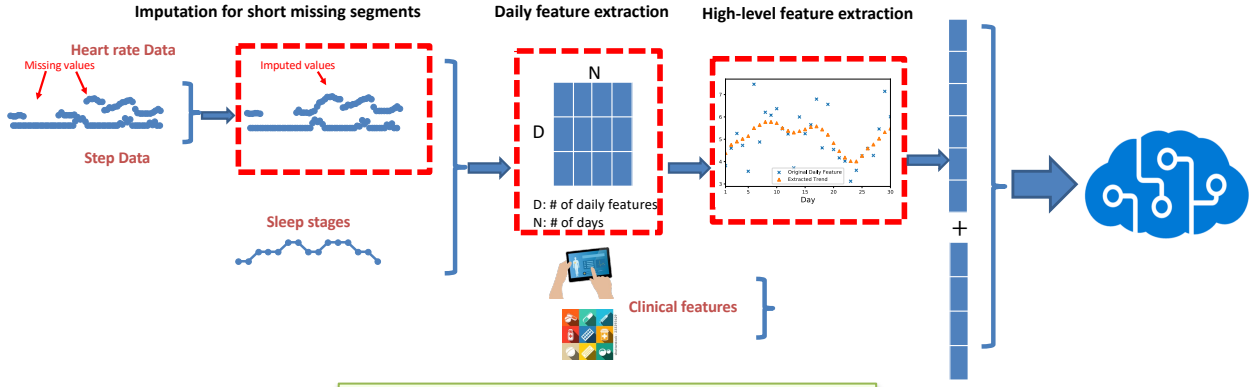


IoMT for Precision Medicine

- **Perioperative care**
  - Surgical complications after pancreatic surgery [Standard of Care in Surgical Prehabilitation]
  - Surgical outcomes of periacetabular osteotomy for hip dysplasia [NIH R01]
  - Recovery outcomes after spine surgery
- **Mental health care**
  - Mental disorders in the community [NIH All of Us]
  - Depression and weight loss of older adults undergoing behavioral therapy [NIH R01]
  - Dynamic cognitive function in youth with diabetes



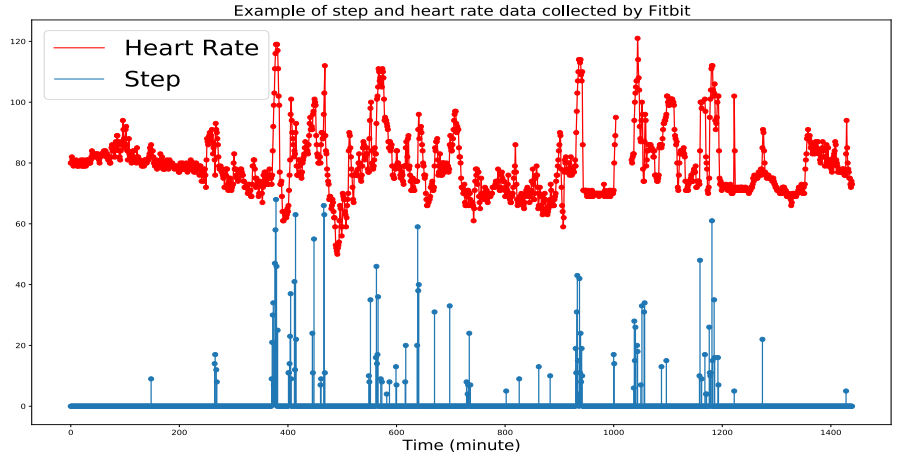
Data Collection

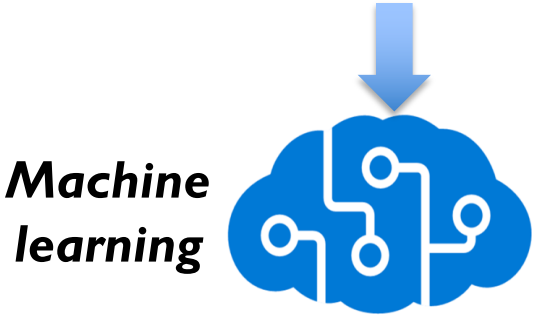


Machine Learning

Wearable Data → Clinical Outcomes

- **Reliable** prediction of **clinical outcomes**
  - Wearable data are **fine-grained, noisy, lossy**
- **Scaling** predictive models
  - Small cohorts: **robust** models with limited data
  - Large population: **deep** learning models
- Prediction of **treatment response**
  - Support personalized **intervention**





Predictions

Clinical decisions and intervention

Predict Clinical Outcomes with Wearables

CPSL Cyber-Physical Systems Laboratory

- Complications after pancreatic surgery
  - **Robust** machine learning based on a **small** patient cohort
- Mental disorders in the community
  - **Deep** learning model for a **large** population
- Treatment response to depression therapy
  - **Personalized** prediction of treatment response

Need: Predict Outcomes of Pancreatic Surgery

- Pancreatic cancer has a 5-year survival rate less than 5%.
- Surgery is the only cure but commonly followed by complications.
- Predict postoperative complications before surgery
  - **Decision support**: suitability for surgery
  - **Intervention**: pre-habilitation

Preoperative Clinic



Surgery



Outpatient Recovery



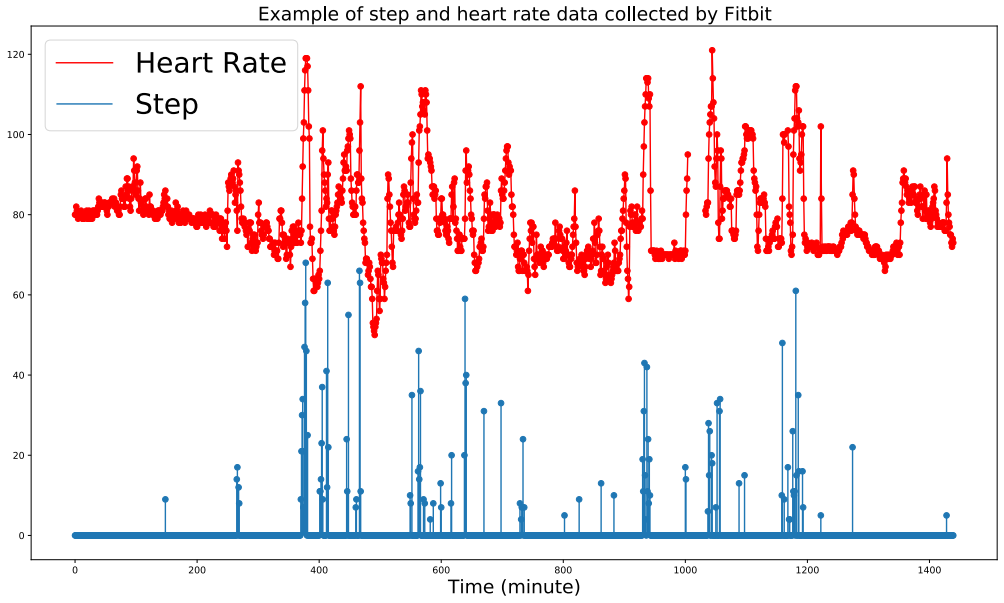
Joint work with Chet Hammill (Surgery), Jingwen Zhang, Dingwen Li, Ruixuan Dai (CSE)

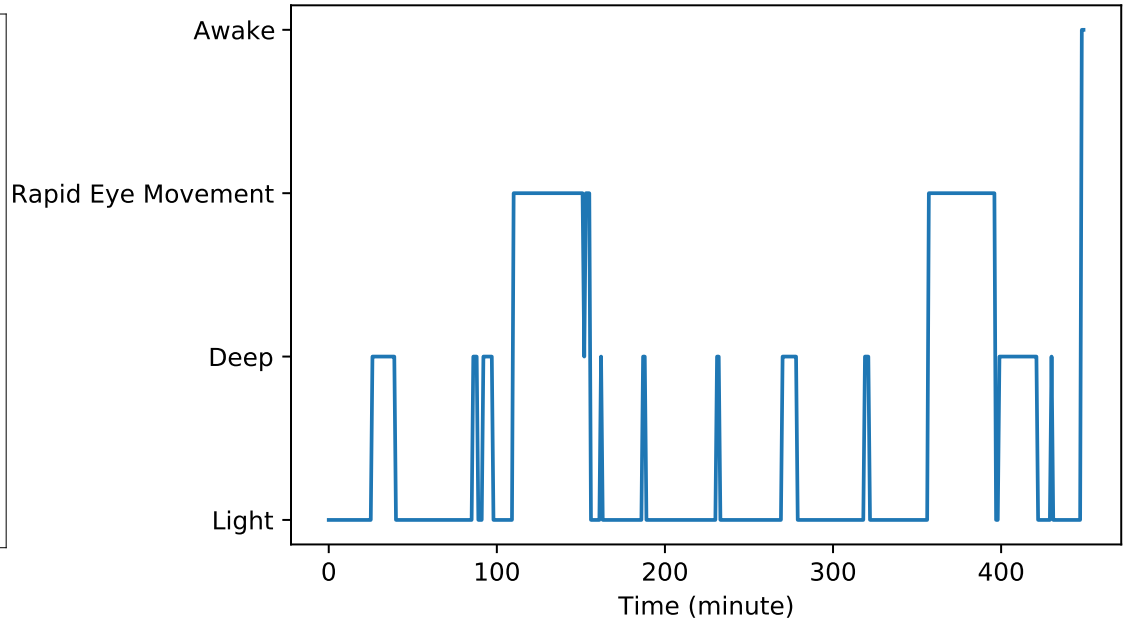
Predict Postoperative Complications

- Goal: predict postoperative complications with preoperative data
  - Wearable data collected by Fitbit wristband
  - Patient clinical characteristics from medical record
- Prospective study: 61 patients undergoing pancreatic surgery
  - 25 (40.98%) experienced complications
- Machine learning approach
  - Small cohort → avoid complex (deep) models
  - Extract features from wearable time series + shallow machine learning models
  - Integrate wearable data and patient clinical characteristics

Fitbit Data

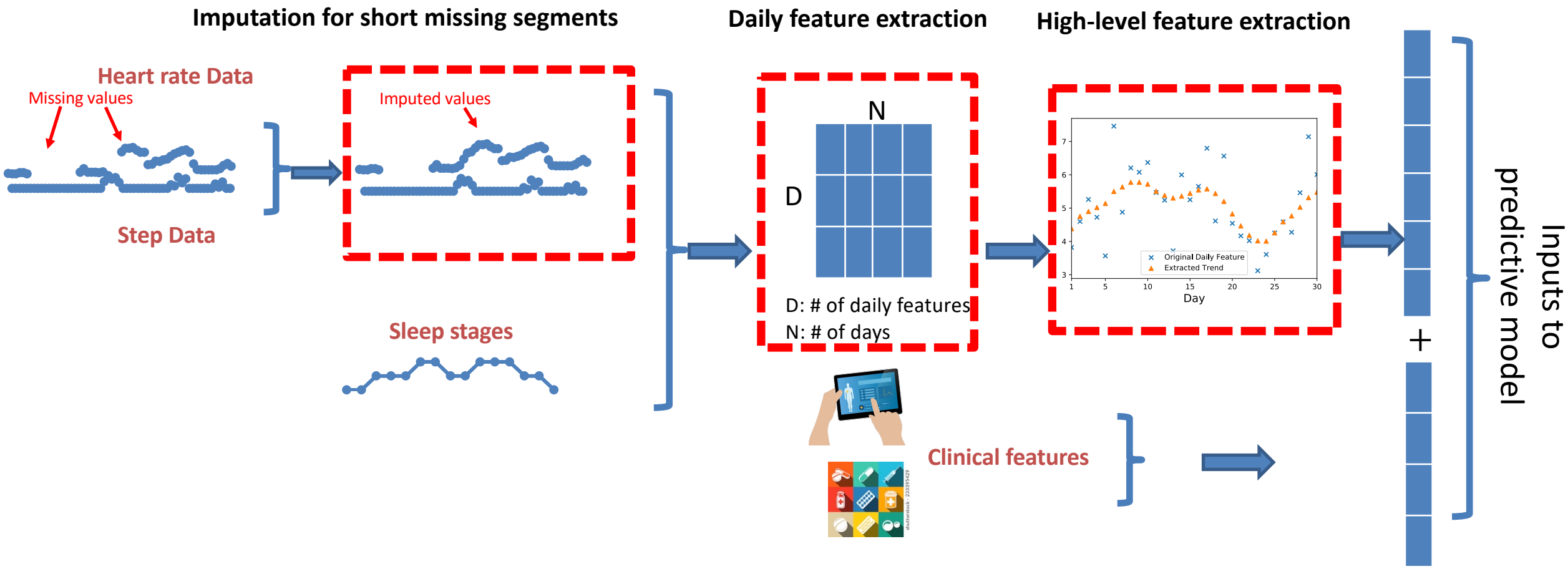
- Wristband provides time series data: step count, heart rate, and sleep stage.





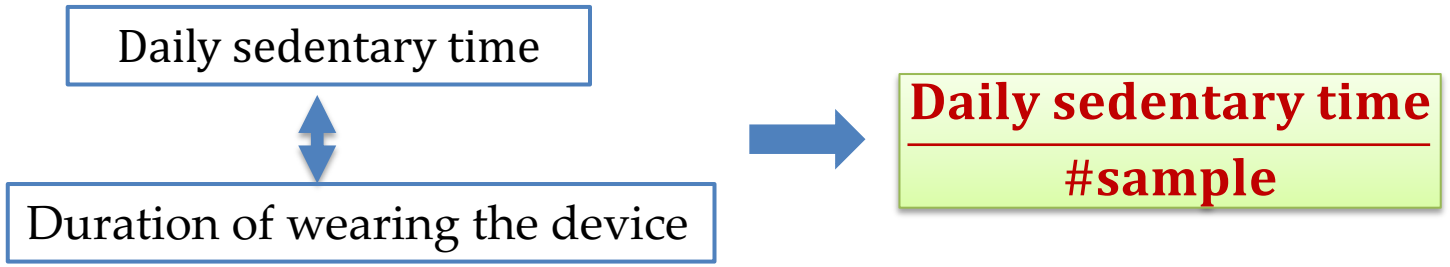
***Fine-grained, lossy, noisy* time-series**

Robust Feature Engineering Pipeline



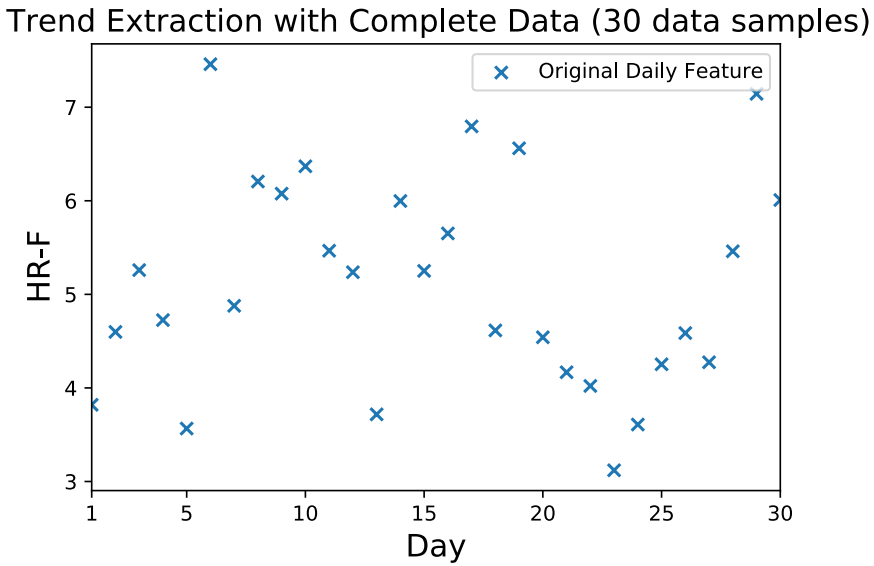
Raw Time Series → Daily Features

- **Statistical** features
  - max, sum, skewness, kurtosis, energy, entropy, inertia…
- **Semantic** features
  - Activity: daily steps, sedentary/active time, active/sedentary bout counts…
  - Sleep: awake counts, sleep efficiency, time of sleep, time before falling asleep…
- **Standardization**: account for missing data



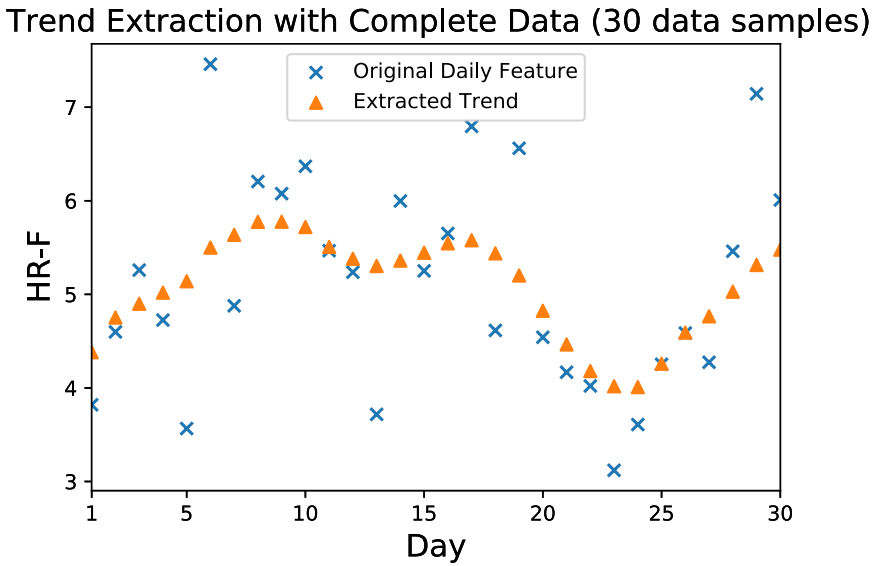
Daily Features → High-level Features

- Singular Spectrum Analysis (SSA)
  - **denoise** time series of daily features



Daily Features → High-level Features

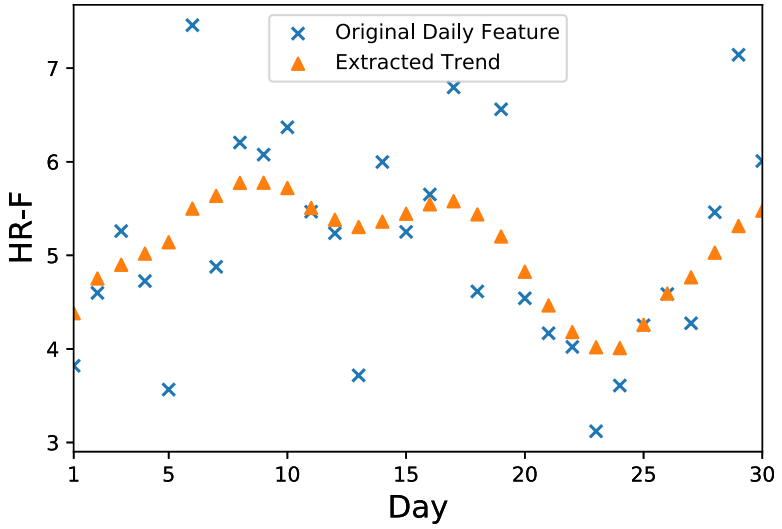
- Singular Spectrum Analysis (SSA)
  - **denoise** time series of daily features



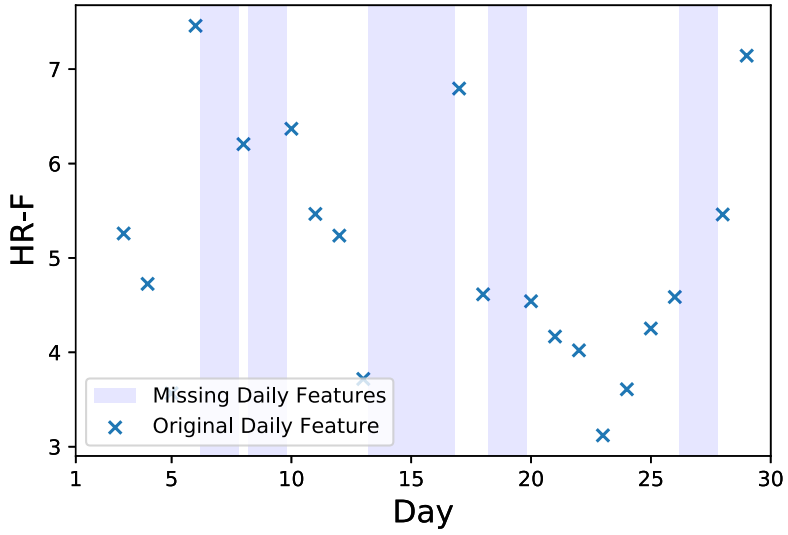
Daily Features → High-level Features

- Singular Spectrum Analysis (SSA)
  - **denoise** time series of daily features.
  - **impute** missing daily features.

Trend Extraction with Complete Data (30 data samples)



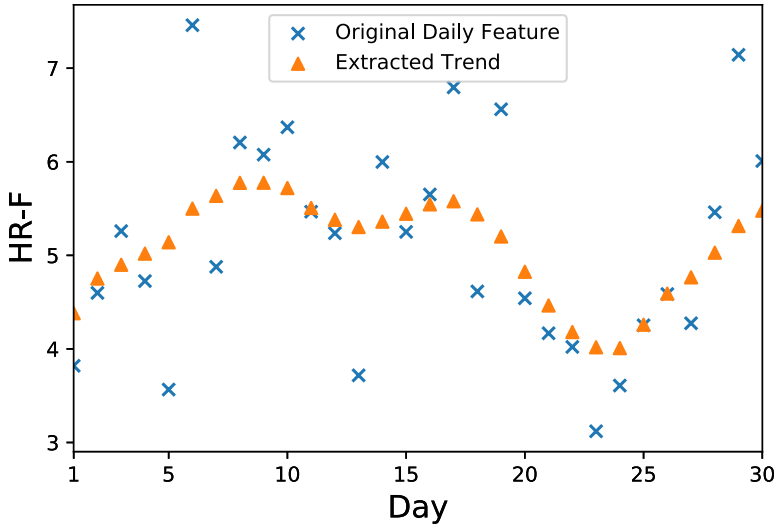
Trend Extraction with Missing Components (20 data samples)



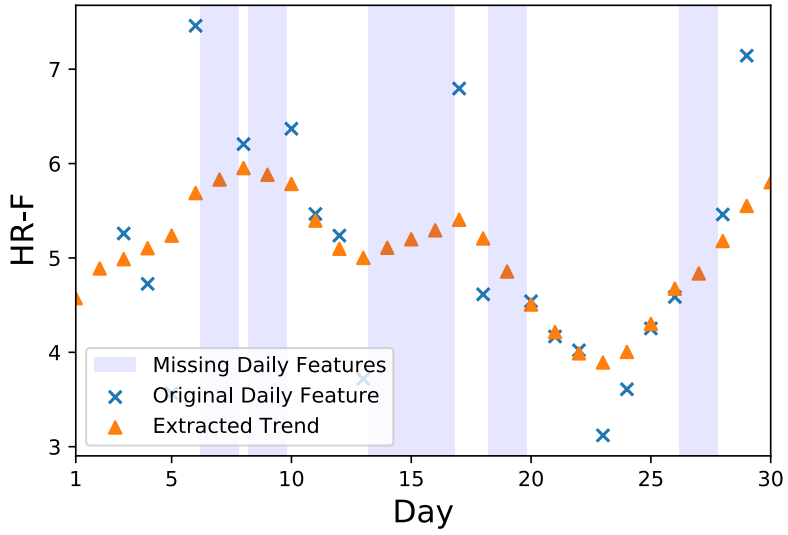
Daily Features → High-level Features

- Singular Spectrum Analysis (SSA)
  - **denoise** time series of daily features.
  - **impute** missing daily features.

Trend Extraction with Complete Data (30 data samples)



Trend Extraction with Missing Components (20 data samples)



- **mean**
- **standard deviation**
- **slope**

Robust Prediction of Surgical Complications

CPSL Cyber-Physical Systems Laboratory

- **Machine learning models outperform 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
- AUROC: Area Under the Receiver Operating Characteristic Curve
- AUPRC: Area Under the Precision-Recall Curve

Robust Prediction of Surgical Complications

CPSL Cyber-Physical Systems Laboratory

- Machine learning models outperform traditional surgical risk scores.
- **Wearable data + clinical characteristics → best predictive performance**

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
- AUROC: Area Under the Receiver Operating Characteristic Curve
- AUPRC: Area Under the Precision-Recall Curve

Handling Missing and Lossy Data

CPSL Cyber-Physical Systems Laboratory

Impact: Standardization of Daily Features

Daily Features	AUROC	AUPRC	Sensitivity	Specificity
Non-standardized	0.6919 (0.0116)	0.6332 (0.0289)	0.4320 (0.0560)	0.8611 (0.0000)
Standardized	**0.7326 (0.0074)**	**0.7192 (0.0154)**	**0.5480 (0.0440)**	0.8583 (0.0083)

→ *Missing samples*

Impact: SSA of Daily Features

Feature Extraction	AUROC	AUPRC	Sensitivity	Specificity
Without SSA Denoising	0.6831 (0.0086)	0.5923 (0.0224)	0.3720 (0.0402)	0.8583 (0.0083)
With SSA	**0.7326 (0.0074)**	**0.7192 (0.0154)**	**0.5480 (0.0440)**	0.8583 (0.0083)

→ *Noisy, missing daily features*

Summary: Predict Postoperative Complications

- Extract **high-level** features from **fine-grained** wearable data
  - Time series data (minute) → Daily features
  - Daily features time series → High-level features

- Extract **robust** features from **noisy and lossy** wearable data
  - Standardized daily features from incomplete time series
  - Singular spectrum analysis for high-level feature extraction

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 (UbiComp'22), 6(2), Article 87, July 2022.

Predict Clinical Outcomes with Wearables

- Complications after pancreatic surgery
  - **Robust** machine learning based on a **small** patient cohort
- Mental disorders in the community
  - **Deep** learning model for a **large** population
- Treatment response to depression therapy
  - **Personalized** prediction of treatment response

Mental Health Crisis

- Mental disorders are prevalent.
  - **~3.8%** of the population (i.e., **280 million**) experience depression (WHO).
- Over **50%** of patients are not recognized or treated.
- Clinical visit is time-consuming and expensive.
  - Hindering timely diagnosis and intervention



- Detect mental disorders with wearables devices?
  - Unobtrusive, multi-modal sensing
  - Activities, heart rate, and sleep are associated with mental health



Need: Large and Diverse Cohorts

- Wearable mental health studies: **small** cohorts with **limited diversity**
  - 18 pregnant women [UbiComp'19]
  - 48-652 college students [UbiComp'14, UbiComp'19, J. Biomed. Inform. '22]
  - 1,002 healthy subjects [NPJ Digit. Med. '18]

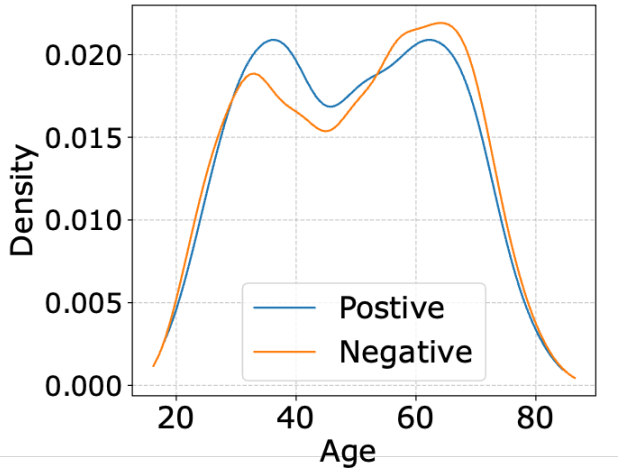
- Machine learning approaches
  - Shallow models rely on ad hoc feature engineering.
  - Deep models tend to overfit.

- **Small datasets limit the rigor and generality of results!**

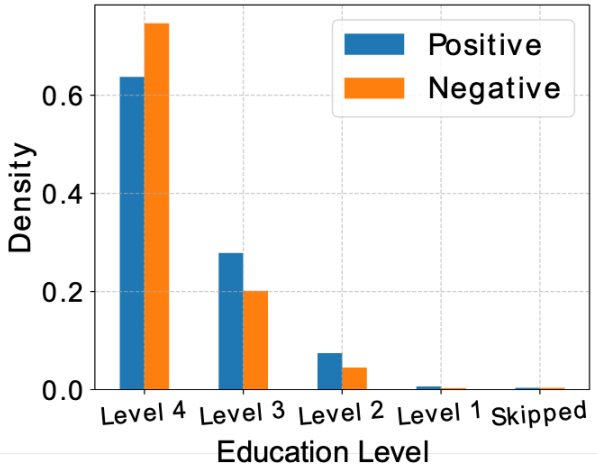
Deep Learning on a Large and Diverse Dataset

- **8,996** participants of the NIH All of Us program
- Including **1,247** with mental (depressive/anxiety) disorders
- **Longitudinal wearable data** and **mental health diagnosis**

Age: mean 48.6 (s.d. 15.9)



Education



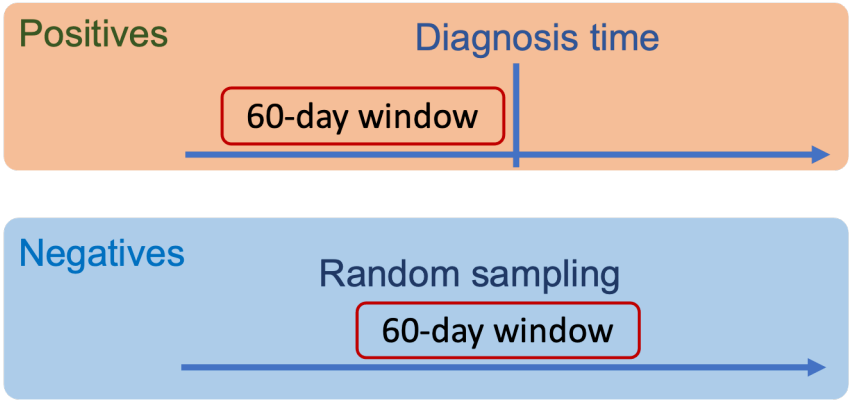
Level1: below high school
Level2: twelve Or GED
Level3: College One to Three
Level4: College graduate or above

- **WearNet**: deep model for detecting mental disorders with wearables

Joint work with Thomas Kannampallil (Informatics), Laura Jean Bierut (Psychiatry), Ruixuan Dai (CSE)

Outcomes and Inputs

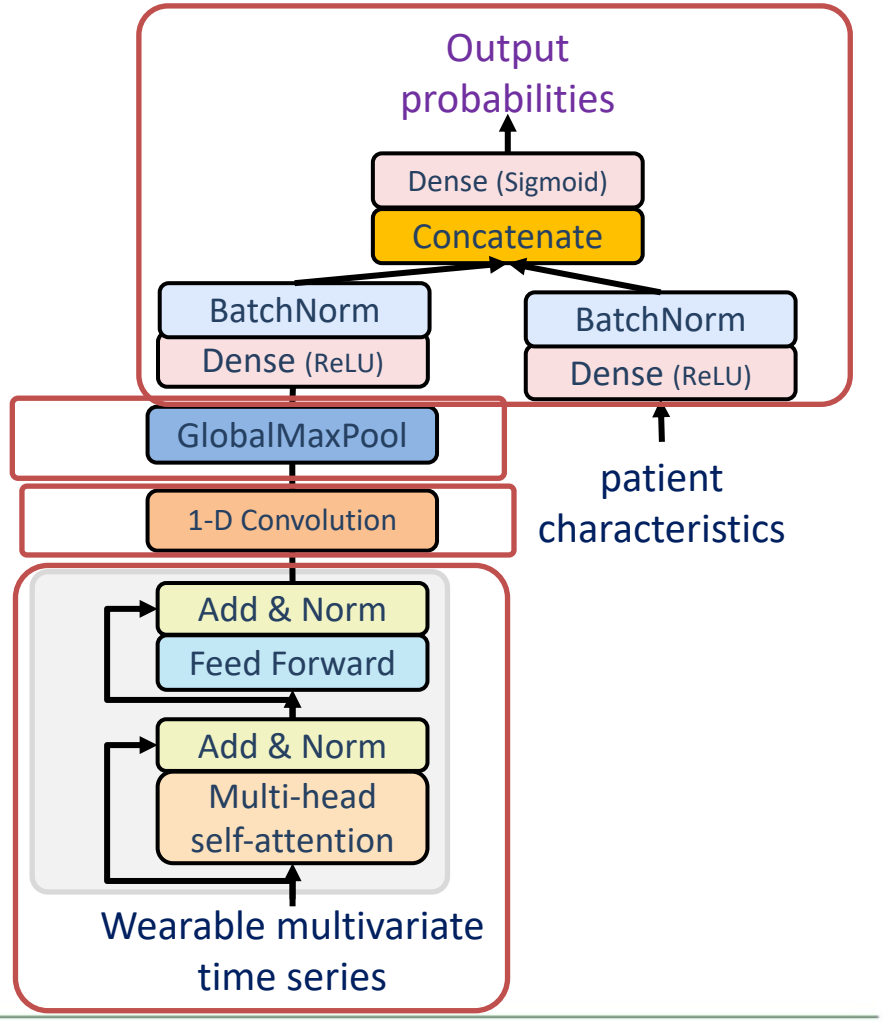
- Mental disorders
  - 20 **depressive/anxiety** disorder diagnoses identified by clinical experts
  - Positive label: a participant with any of the diagnoses



- Input features
  - Wearable data in a 60-day window
    - 10 daily features from Fitbit cloud
    - Derived based on step and heart rate time series
    - e.g., total steps, average heart rate, sedentary minutes…
  - Static data from electrical health records and surveys
    - Age, race, ethnicity, gender, education, smoke history, alcohol history

WearNet: Deep Model for Detecting Mental Disorders

- Combining transformer encoder and convolutional neural network
- Transformer encoder (multi-head self-attention)
  - Identifies patterns across multiple timestamps
- Convolutional neural network (1-d)
  - Integrates neighborhood patterns
- Global max pooling
  - Identifies one global pattern for robustness
- Integrate wearable and static data at the top
  - Captures the underlying characteristics



Detection Performance

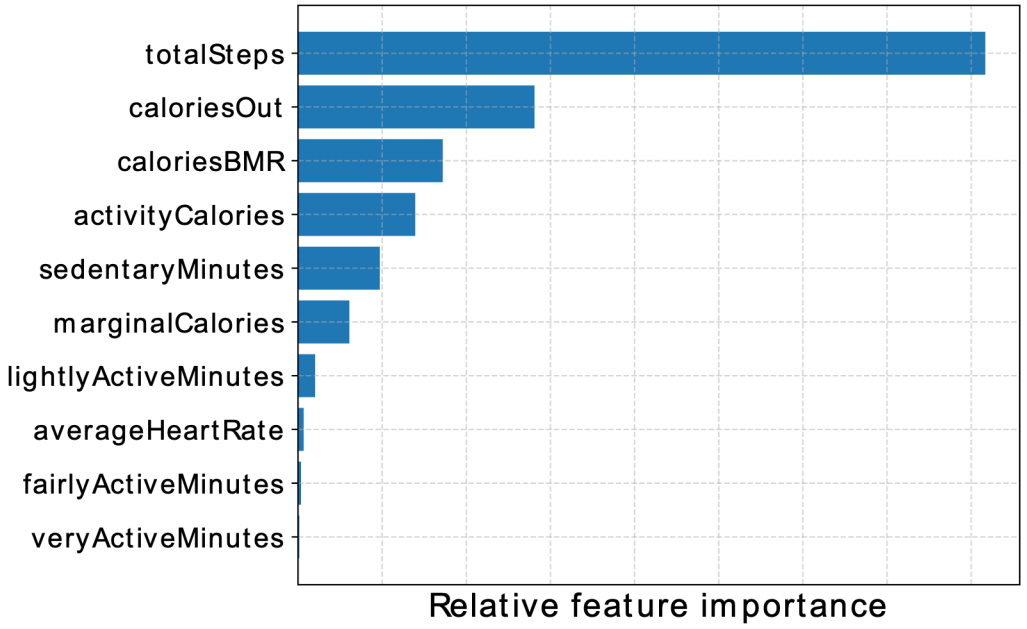
- **AUROC**: Area Under the Receiver Operating Characteristic Curve
- **AUPRC:** Area Under the Precision-Recall Curve

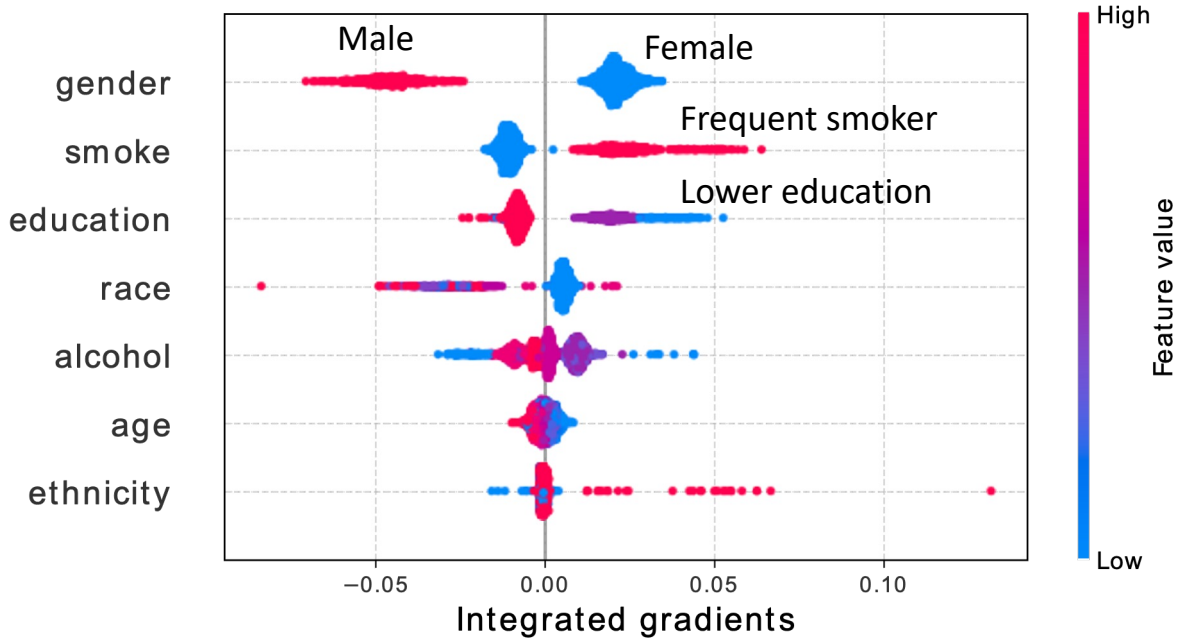
Category	Model	AUROC	AUPRC	
Shallow Models	LR	0.701(0.000)	0.351(0.000)	
	SVM	0.592(0.000)	0.290(0.000)	
	RF	0.661(0.005)	0.349(0.007)	
	GBDT	0.685(0.001)	0.365(0.000)	
Deep Models	bi-LSTM	0.702(0.015)	0.464(0.011)	**RNN and TCN Models**
	BRITS	0.693(0.012)	0.445(0.011)	
	CrossNet	0.682(0.021)	0.429(0.014)	
	TCN	0.629(0.021)	0.235(0.024)	
	Informer	0.705(0.008)	0.428(0.011)	**Transformer Models**
	WearNet	**0.717(0.009)**	**0.487(0.008)**	

- **Shallow models** with feature engineering are less predictive.
- **RNN and TCN models** underperform Transformer models.
- **WearNet** achieves the best predictive performance.

Model Explanation

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





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

→ **Match literature**

Summary: Detect Mental Disorders

- Mental disorders are **prevalent** but significantly **underdiagnosed**.
- A **large** and **diverse** cohort with 8,996 participants from All of Us.
- **WearNet**: a deep model for detecting mental disorders with wearables.
- An **unobtrusive** approach to detect mental disorders in the community.

R. Dai, T Kannampallil, S. Kim, V. Thornton, L. Bierut, C. Lu, Detecting Mental Disorders with Wearables: A Large Cohort Study, ACM/IEEE Conference on Internet of Things Design and Implementation (IoTDI'23), May 2023.

Predict Clinical Outcomes with Wearables

CPSL Cyber-Physical Systems Laboratory

- Complications after pancreatic surgery
  - **Robust** machine learning based on a **small** patient cohort
- Mental disorders in the community
  - **Deep** learning model for a **large** population
- Treatment response to depression therapy
  - **Personalized** prediction of treatment response

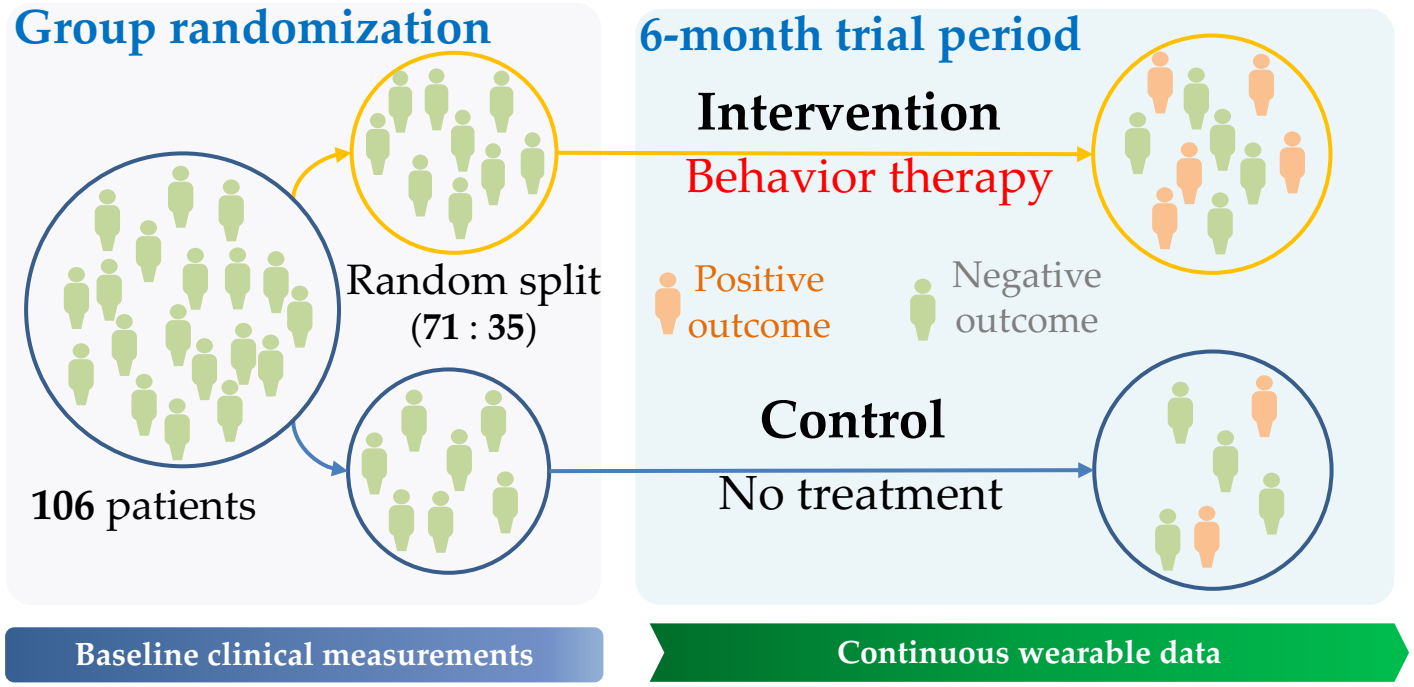
Personalized Prediction of Treatment Response

- Statistical analysis → **population-level** effectiveness of treatment

- **Personalized** prediction → **precision** medicine

- **Machine learning from RCT data**
  - Clinical (baseline): age, anxiety…
  - Fitbit (2 months): heart rate, sleep
  - Depression outcome (at 6 month)

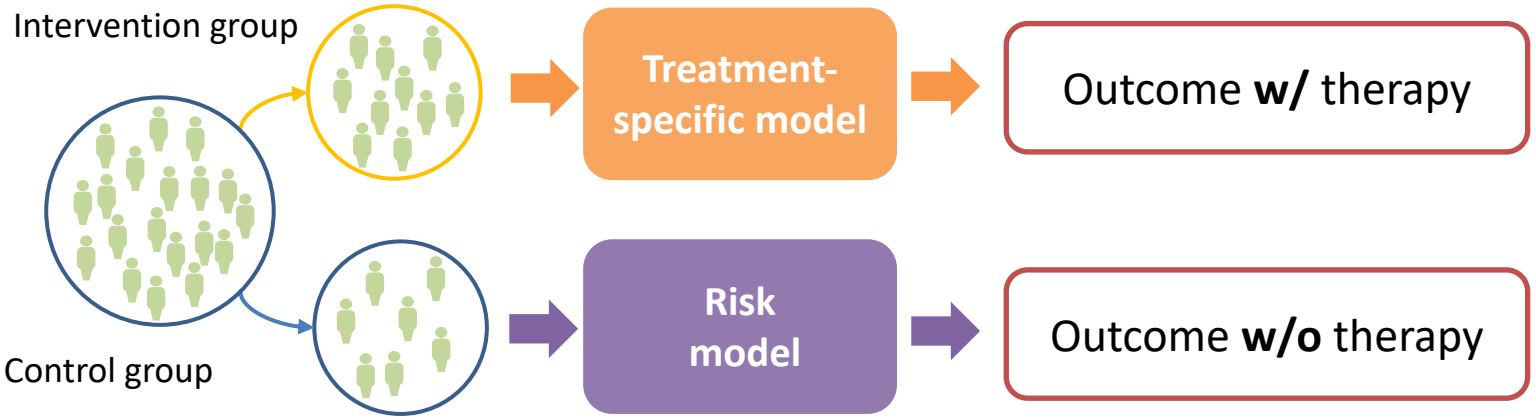
Randomized Controlled Trial of Depression Therapy



Joint work with Thomas Kannampallil (Informatics), Jun Ma (Medicine, UIC), Ruixuan Dai, Jingwen Zhang (CSE)

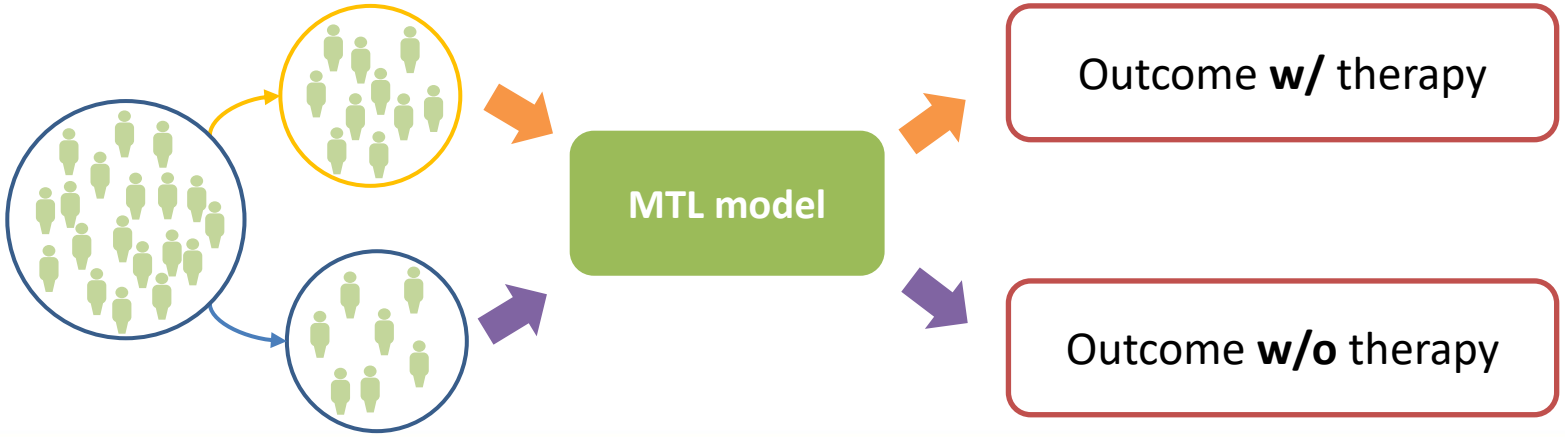
Predicting Personalized Treatment Response

- **Group-specific** models



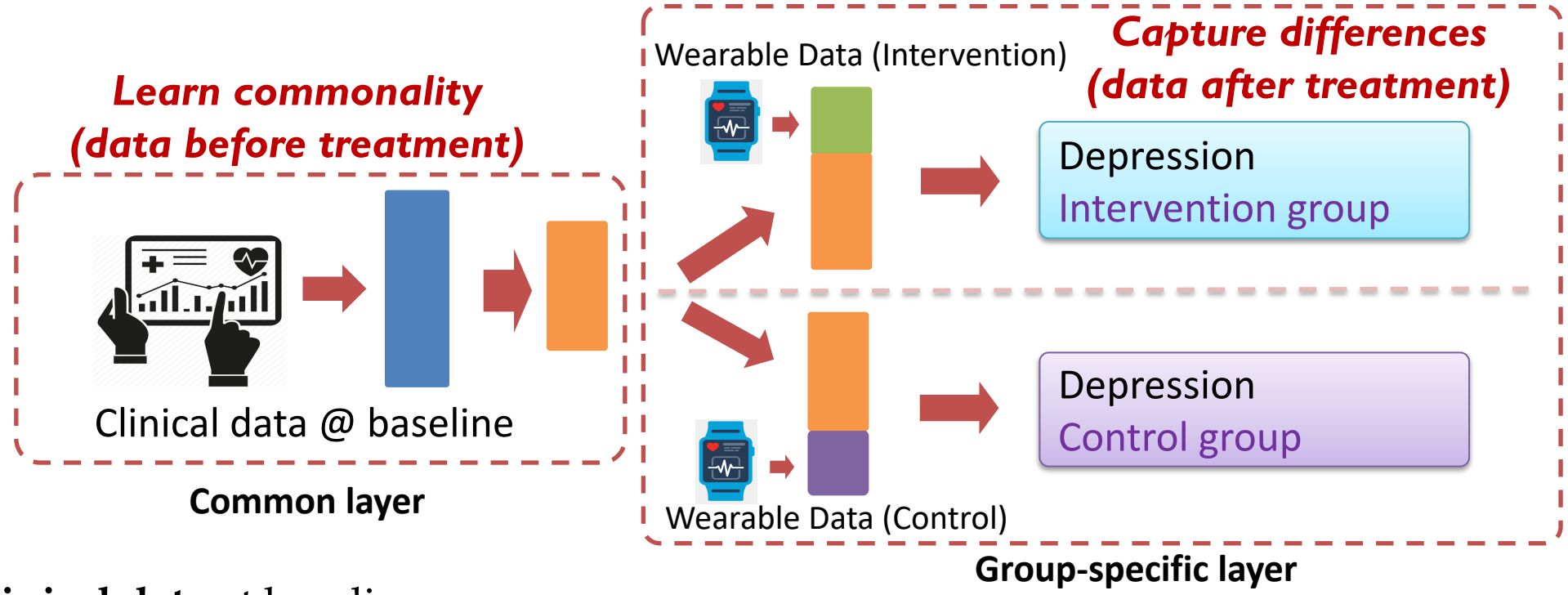
- *Small groups → overfitting*

- **Multi-task learning (MTL)** with a **unified** model



- *Exploit similarity*
- *Capture differences*

Multi-Task Learning Architecture



- **Clinical data** at baseline
  - Statistically similar between groups at baseline
- **Wearable data** during the first 2 months
  - Data of the intervention group reflected effects of therapy

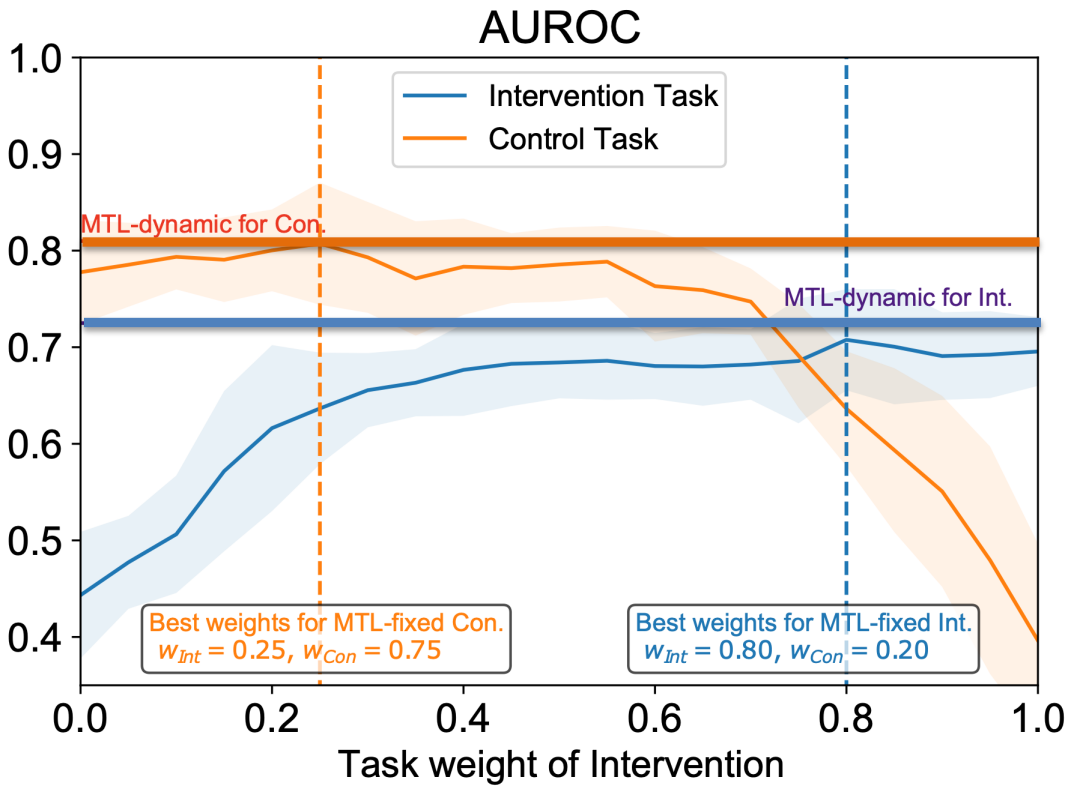
Training Multi-Task Models

- Optimize the loss function for both groups

$$Loss = \boldsymbol{w_{int}} \cdot Loss_{int} + \boldsymbol{w_{con}} \cdot Loss_{con}$$

- Traditional: **Tune** weights → task contributions
  - $w_{int} + w_{con} = 1$
  - Hyperparameters (fixed for the model)
  - Cannot achieve best performance for both groups

- New: **Learn** weights based on task uncertainties
  - Larger uncertainty → lower contribution
  - Optimize both groups automatically



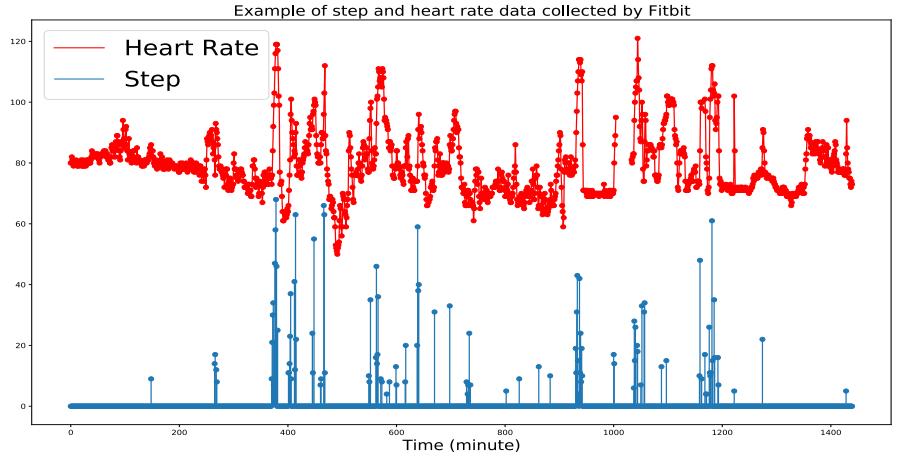
Summary: Predict Treatment Response

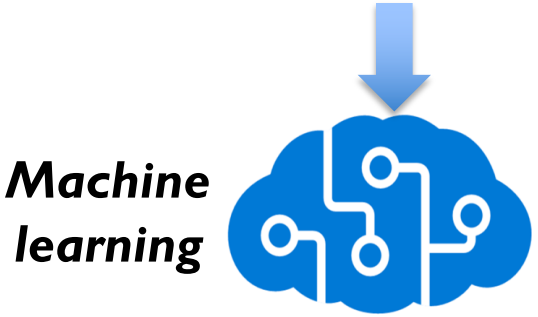
- **Personalized** predictions of **treatment response** → precision medicine
- Learn from data collected in **randomized controlled trials**
- Multi-task learning exploits the **commonalities** while capturing the **differences** between intervention and control groups

R. Dai, T. Kannampallil, J. Zhang, N. Lv, J. Ma, and C. Lu, Multi-Task Learning for Randomized Controlled Trials: A Case Study on Predicting Depression with Wearable Data, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UbiComp'22), 6(2), Article 50, July 2022.

IoMT for Precision Medicine

- **Reliable** prediction of **clinical outcomes**
  - Wearable data are **fine-grained, noisy, lossy**
- **Scaling** predictive models
  - Small cohorts: **robust** models with limited data
  - Large population: exploit **deep** learning models
- Prediction of **treatment response**
  - Support personalized **intervention**





Predictions

Clinical decisions and intervention

Embed IoMT in Healthcare

- Significant **clinical information** can be learned from wearables.
- **AI** is key for extracting clinical information from wearable data.
- Rigorous **clinical studies** are needed to validate IoMT models.
- Precision medicine: **personalized prediction** → **intervention** → **outcome**
- Close **collaboration** between AI and clinical researchers is essential

AIM Research: https://www.cse.wustl.edu/~lu/iomt.html