Internet of Medical Things
Predicting Clinical Outcomes with Wearables

Chenyang Lu
Department of Computer Science & Engineering
Institute of Artificial Intelligence and Internet of Medical Things for Healthcare
Cyber-Physical Systems Laboratory
https://www.cse.wustl.edu/~lu/
Internet of Medical Things

- **Wearables**: wristbands, smartwatches...
  - 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

**A powerful tool for healthcare!**
Wearable Data

- Commonly available: step, heart rate, sleep stages.
- More sensing modalities
  - Oxygen saturation (SpO2)
  - On-wrist skin temperature
  - Breathing rate (during sleep)
  - Heart rate variability
  - ECG (AFib)
  - Stress
- 500+ million wearables sold in 2021

Unprecedented capability of longitudinal monitoring outside hospitals!
Wearable Data → Clinical Outcomes

- **Reliable** prediction of clinical outcomes
  - Wearable data are **fine-grained, noisy, lossy**

- **Personalized** prediction of treatment response
  - Support clinical decisions on **intervention**

- **Scaling** predictive models
  - Small cohorts: **robust** models with limited data
  - Large population: exploit **deep** learning models
Predict Clinical Outcomes with Wearables

- Postoperative complications of patients undergoing surgery
  - Robust machine learning based on a small patient cohort

- Depression remission of older adults undergoing behavioral therapy
  - Personalized prediction of treatment response

- Mental disorders in the community
  - Deep learning model for a large population
Predict Clinical Outcomes with Wearables

- Postoperative complications of patients undergoing surgery
  - Robust machine learning based on a small patient cohort

- Depression remission of older adults undergoing behavioral therapy
  - Personalized prediction of treatment response

- Mental disorders in the community
  - Deep learning model for a large population
Predicting Pancreatic Surgery Outcome

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

- Pancreatic surgery is the only cure but is commonly followed by complications.

- Predict postoperative complications before surgery
  - **Decision support**: suitability for surgery
  - **Intervention**: pre-habilitation

Joint work with Chet Hammill (Surgery), Jingwen Zhang, Dingwen Li, Ruixuan Dai (CSE)
Goal: predict **postoperative** complications with **preoperative** data
- Wearable data collected by Fitbit (time series)
- Patient clinical characteristics from medical record (static data)

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

Heart rate Data
Step Data
Sleep stages

Imputation for short missing segments

Daily feature extraction
D: # of daily features
N: # of days

High-level feature extraction

Clinical features
Inputs to predictive model

Missing values
Daily Features Extraction

- Daily features are generated from step, heart rate, and sleep time series.
- Statistical features
  - First-order features: max, sum, skewness, kurtosis
  - Second-order features: energy, entropy, inertia, local homogeneity
- Detrended fluctuation analysis: capture local fluctuation
- Semantic features: sedentary/active information

- **Standardization** is used to accommodate missing data

\[
\text{Daily sedentary time} = \frac{\text{Duration of wearing the device}}{\# \text{ Data samples}}
\]
Robust Feature Engineering Pipeline

Imputation for short missing segments

Daily feature extraction

High-level feature extraction

Inputs to predictive model

Heart rate Data

Step Data

Sleep stages

Missing values

Imputed values

D: # of daily features

N: # of days

Clinical features

Imputation for short missing segments

Daily feature extraction

High-level feature extraction

Inputs to predictive model

Heart rate Data

Step Data

Sleep stages

Missing values

Imputed values

D: # of daily features

N: # of days

Clinical features

Inputs to predictive model
High-level Features Extraction

- Singular Spectrum Analysis (SSA)
  - **denoise** time series of daily features.
High-level Features Extraction

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

Trend Extraction with Complete Data (30 data samples)
High-level Features Extraction

- Singular Spectrum Analysis (SSA)
  - denoise time series of daily features.
  - robust to missing daily features.

Trend Extraction with Complete Data (30 data samples)  Trend Extraction with Missing Components (20 data samples)
High-level Features Extraction

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

- Machine learning models outperform standard surgical risk scores.
  - x2 AUPRC
  - x3 sensitivity at the same specificity

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

- 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

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

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

- 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
Impact: Standardization of Daily Features

<table>
<thead>
<tr>
<th>Daily Features</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-standardized</td>
<td>0.6919 (0.0116)</td>
<td>0.6332 (0.0289)</td>
<td>0.4320 (0.0560)</td>
<td>0.8611 (0.0000)</td>
</tr>
<tr>
<td>Standardized</td>
<td>0.7326 (0.0074)</td>
<td>0.7192 (0.0154)</td>
<td>0.5480 (0.0440)</td>
<td>0.8583 (0.0083)</td>
</tr>
</tbody>
</table>

Impact: SSA of daily features

<table>
<thead>
<tr>
<th>Feature Extraction</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without SSA Denoising</td>
<td>0.6831 (0.0086)</td>
<td>0.5923 (0.0224)</td>
<td>0.3720 (0.0402)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>With SSA</td>
<td>0.7326 (0.0074)</td>
<td>0.7192 (0.0154)</td>
<td>0.5480 (0.0440)</td>
<td>0.8583 (0.0083)</td>
</tr>
</tbody>
</table>

Handle missing data

Handle noisy and missing data
Summary: Predict Clinical Outcomes

- Extract **high-level** features from fine-grained wearable data
  - Time series data (minute) $\rightarrow$ Daily features (day)
  - Daily features (day) $\rightarrow$ High-level features (longitudinal monitoring)

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

---

Postoperative complications of patients undergoing surgery
  - Robust machine learning based on a small patient cohort

Depression remission of older adults undergoing behavioral therapy
  - Personalized prediction of treatment response

Mental disorders in the community
  - Deep learning model for a large population
Randomized Controlled Trial: Depression Therapy

- Statistical analysis → population-level effectiveness of treatment

- Personalized prediction → precision medicine

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

---

Group randomization

106 patients

Random split (71 : 35)

Intervention

Behavior therapy

- Positive outcome
- Negative outcome

Control

No treatment

6-month trial period

Baseline clinical measurements

Continuous wearable data

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

- **Group-specific models**
  - Intervention group
  - Treat-specific modelling
  - Outcome w/ therapy
  - Control group
  - Risk modelling
  - Outcome w/o therapy

  - Small groups ↠ overfitting

- **Multi-task learning (MTL) with a unified model**
  - Intervention group
  - MTL modelling
  - Outcome w/ therapy
  - Control group
  - MTL modelling
  - Outcome w/o therapy

  - Enlarge training dataset
  - Exploit similarity between groups
Multi-Task Learning Architecture

- **Clinical data** at baseline
  - Statistically similar between 2 groups at baseline

- **Wearable data** during the first 2 months
  - Data of the intervention group reflected effects of therapy
Training MTL Models

- Optimize the loss function for both groups
  \[ \text{Loss} = w_{\text{int}} \cdot \text{Loss}_{\text{int}} + w_{\text{con}} \cdot \text{Loss}_{\text{con}} \]

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

- New: **Learn** weights based on task uncertainties in the training process
  - Larger uncertainty \( \rightarrow \) lower contribution
  - Optimize both groups automatically

![Graph showing AUROC for intervention and control tasks with task weights](image)
Evaluation

- **Group-specific**: separate model for each group
  - **MTL-1**: same MTL model but trained on a single group

- **MTL on combined groups**
  - **MTL-fixed**: with fixed weights
  - **MTL-dynamic**: with dynamic weights

<table>
<thead>
<tr>
<th>Category</th>
<th>Model</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>group-specific</td>
<td>Logistic Regression</td>
<td>0.697(0.050)</td>
<td>0.794(0.067)</td>
</tr>
<tr>
<td></td>
<td>MTL-1</td>
<td>0.695(0.032)</td>
<td>0.784(0.049)</td>
</tr>
<tr>
<td>MTL</td>
<td>MTL-fixed</td>
<td>0.707(0.052)</td>
<td>0.807(0.063)</td>
</tr>
<tr>
<td></td>
<td>MTL-dynamic</td>
<td>0.725(0.059)</td>
<td>0.813(0.077)</td>
</tr>
</tbody>
</table>

- **MTL models trained on combined groups improve performance**
- **MTL-dynamic optimizes performance for both groups**
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

---

Predict Clinical Outcomes with Wearables

- Postoperative complications of patients undergoing surgery
  - Robust machine learning based on a small patient cohort

- Depression remission of older adults undergoing behavioral therapy
  - Personalized prediction of treatment response

- Mental disorders in the community
  - Deep learning model for a large population
Mental disorders are prevalent
- Over 50% of patients are not recognized or treated.

Previous wearable studies involved small cohorts
- Relied on feature engineering + shallow machine learning models

Can IoMT detect mental disorders in the community?
- Exploit end-to-end deep learning on a large dataset
Detect mental disorder (depression & anxiety) using
- **wearable data**: daily features
- **static data**: age, race, ethnicity, gender, education, smoke, alcohol

All of Us: **8,996** participants with wearables, **1,247** with mental disorder diagnoses

**WearNet**: end-to-end deep model
- Learn directly from raw daily features
- No feature engineering
- Integrate wearable and static data

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

- **Reliable** prediction of clinical outcomes
  - Wearable data → clinical outcomes

- **Personalized** prediction of treatment response
  - Randomized controlled trials → personalized intervention

- **Scaling** predictive models
  - Small patient cohorts → robust feature engineering
  - Large population → deep learning models

Unleash the power of wearables for healthcare!
Embed IoMT in Healthcare

➢ It is all about **intervention** and improving outcomes
  - Predicting outcomes is only the first step
  - Prediction → intervention → outcome

➢ Internet of **Medical** Things
  - Close collaboration with clinicians and healthcare organizations
  - Clinical problems and expertise are essential drivers of the field

➢ IoMT + Informatics + Clinicians → Healthcare **Delivery**
  - Integration with informatics infrastructure and clinical workflow
  - Make IoMT work for clinicians (human-in-the-loop AI, explainable AI…)

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