Internet of Medical Things

Predicting Clinical Outcomes with Wearables

Chenyang Lu
Cyber-Physical Systems Laboratory
Department of Computer Science & Engineering

https://www.cse.wustl.edu/~lu/
Internet of Medical Things

- **Wearables**: wristbands, smartwatches...
  - Continuous monitoring: activity, heart rate, sleep, location...

- **Connectivity**: Bluetooth, WiFi, cellular...
  - Real-time monitoring and just-in-time intervention

- **Cloud**: computing and storage.
  - Scalable to large cohorts

- **Analytics**: machine learning and signal processing
  - Interpret data and predict outcomes

*A powerful tool for healthcare inside and outside hospitals!*
Wearables as Healthcare Tools

Open, programmable platform
Wear OS, Research Kit, onboard analytics

Continuous, passive measurements
activity, heart rate, sleep, location...

Two-way communication
ecological momentary assessments
Increasing Sensing Capabilities

- Commonly available: step, heart rate, sleep stages.
- New: Fitbit Sense
  - Oxygen saturation (SpO2)
  - On-wrist skin temperature
  - Breathing rate (during sleep)
  - Heart rate variability
  - ECG (AFib)
  - Stress
Wearable Use Cases

- Predicting clinical outcomes
  - Readmission and mortality of congestive heart failure patients
  - Post-operative complications and readmission of surgical patients

- Measuring mobility
  - Timed Up and Go with smartwatches

- Measuring mental health
  - Ecological momentary assessment (EMA)
  - Mental health model based on physiological signals
Interdisciplinary research in close collaboration with clinicians
- Complications and readmissions of patients undergoing pancreas surgery
- Clinical deterioration and readmissions of congestive heart failure patients
- Impact of glucose variability on cognitive function in youth with diabetes
- Predict recovery after spine surgery
- Mental health of older adults undergoing digital behavioral health interventions

Robust pipelines for **data collection** and **feature engineering** for Fitbit

**Predict Clinical Outcomes with Fitbit**

- Data Collection
- Feature Engineering

- Smart Phone → Fitbit Cloud → Poll request → Transfer data
- Imputation with KNN
  - Heart rate
  - Step
- Daily feature extraction
- High-level feature extraction
- Inputs to predictive model
  - HR
  - Step
  - Imputed value
  - Sleep stages
  - Comorbidities
  - One-hot encoding
  - Predict clinical outcomes
Surgical Complications and Readmissions

- Patients undergoing pancreatectomy are monitored with Fitbit before surgery, during hospital stay, and 30 days post-discharge.

- Machine learning models predict complications and readmissions.
- 62 patients enrolled (goal: 130).

Joint work with Chet Hammill (Surgery), Jingwen Zhang, Dingwen Li, Ruixuan Dai (CSE)
Readmissions of Heart Failure Patients

- Hospital readmission rate is high for congestive heart failure patients.
- Predict deterioration (readmission and death) after discharge
  - Fitbit provides continuous monitoring of outpatients
  - Machine learning based on Fitbit data

Joint work with Thomas Bailey (Infectious Diseases), Marin Kollef (Critical Care), Dingwen Li (CSE)
Data Collection Infrastructure

- Mature data collection and storage pipeline
- Used in multiple studies

Diagram:
- Fitbit
- Smartphone
- Fitbit Cloud
- Database
- Transfer data
- Poll request

10/21/2021
Chenyang Lu
Protocol: Heart Failure Patients

- 25 heart failure patients were recruited.
  - Each patient is given Fitbit ChargeHR wristband.
  - Continuously monitor patients after discharge.
  - Outcomes: 60-day deterioration events.

- Heart rate (HR), step count and sleep stages were collected.
  - Sampling period: 1 min (step, heart rate); 1 day (sleep)
Yield

- **Yield**: fraction of samples successfully collected and stored in database
- Participants with data yield > 80%
  - Step: 88% of participants
  - Heart rate: 60% of participants
  - Sleep: 68% of participants

![Yield per participant chart]

Chenyang Lu
Compliance

- Participants with data yield > 80%
  - Step: 88% of participants
  - Heart rate: 60% of participants → wore Fitbit properly
  - Sleep: 68% of participants → wore Fitbit at night

Feasibility of long-term, continuous monitoring

Yield per participant

Chenyang Lu 10/21/2021
Latency

- Median: 8.6 min; 99% percentile: 22.5 hour
  - Did not cause data loss, as Fitbit can locally store data for 7 days.
- **Feasibility for daily intervention**
  - Can be reduced if analysis performed natively in Fitbit cloud
Machine Learning for Wearables

- Learning from wearable data
  - Predict clinical outcomes
  - Discover risk factors associated with outcome
  - Support clinical decisions
  - Enable early intervention to improve outcome

- Wearable data: fine-grained, noisy, lossy
  - Sampling period: 1 min (step, heart rate); 1 day (sleep)
  - Small patient cohorts → avoid deep models

- How to extract robust and predictive features from wearable data?
Feature Engineering Pipeline

- Extract features as input to shallow machine learning models
  - Imputation for missing data
  - Extract predictive features from wearable data
  - Integration with clinical data

- Imputation with two-dimensional KNN
- Daily feature extraction
- High-level feature extraction

Heart rate
Step
Sleep stages

Inputs to predictive model

Heart rate
Step
Sleep stages

Inputs to predictive model
Two-Level Feature Extraction

**Daily Features**

- **Statistical features**
  - 1st order: mean, max, min, skewness, kurtosis
  - 2nd order: energy, entropy, correlation, inertia and local homogeneity

- **Detrended Fluctuation Analysis**
  - Statistical self-affinity of time series
  - The fluctuation is then used as feature

- **Sedentary behavior**

**High-Level Features**

- **Robust Singular Spectrum Analysis of noisy and incomplete daily features**

![Graph showing trend extraction from incomplete DFA-Sleep-20 Daily Feature](image)
Rich Features of Wearable Data

- Heart rate (HR), step count and sleep quality were collected.
  - Sampling period: 1 min (step, heart rate); 1 day (sleep)

- Statistical features
  - First- and second-order features extracted from sliding window
  - 1\textsuperscript{st} order: mean, max, min, skewness, kurtosis
  - 2\textsuperscript{nd} order: energy, entropy, correlation, inertia and local homogeneity

- Detrended Fluctuation Analysis
  - Determine statistical self-affinity of time series
  - The fluctuation is then used as feature

- Sedentary behavior
Assess Predictability

- Analysis of Variance (ANOVA)
  - Test significant differences in the features between patients of different outcomes (e.g., deteriorated vs. non-deteriorated patients)

- Assess feasibility of learning a stable predictor with the set of features
  - Significant differences in features $\rightarrow$ predictability of outcomes

<table>
<thead>
<tr>
<th>Features</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR skewness</td>
<td>7.9125</td>
<td>0.0099</td>
</tr>
<tr>
<td>HR correlation</td>
<td>5.4789</td>
<td>0.0283</td>
</tr>
<tr>
<td>HR DFA 10</td>
<td>5.3353</td>
<td>0.0302</td>
</tr>
<tr>
<td>Restless duration</td>
<td>5.2912</td>
<td>0.0308</td>
</tr>
<tr>
<td>Time in bed</td>
<td>5.2663</td>
<td>0.0312</td>
</tr>
</tbody>
</table>

Features with the largest F values.

*High F-statistic and low p-value $\rightarrow$ significant difference between group means.*
Feature Selection

- Select features using sequential forward feature selection
  - Avoid overfitting
  - Improve performance

- Features selected by the models have significant differences in ANOVA test

<table>
<thead>
<tr>
<th>Features</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>sleep DFA 60</td>
<td>0.2254</td>
<td>0.6395</td>
</tr>
<tr>
<td>min asleep</td>
<td>4.3128</td>
<td>0.0492</td>
</tr>
<tr>
<td>daily step</td>
<td>4.3625</td>
<td>0.0480</td>
</tr>
<tr>
<td>restless count</td>
<td>4.2324</td>
<td>0.0512</td>
</tr>
<tr>
<td>awake count</td>
<td>4.0429</td>
<td>0.0562</td>
</tr>
<tr>
<td>min awake</td>
<td>2.2073</td>
<td>0.1509</td>
</tr>
<tr>
<td>HR LH</td>
<td>4.0282</td>
<td>0.0566</td>
</tr>
<tr>
<td>HR DFA 10</td>
<td>5.3353</td>
<td>0.0302</td>
</tr>
</tbody>
</table>
Evaluation with Small Data

- **Leave-one-out cross validation**
  - Leave a patient out for testing and train the model with rest of patients.
  - Iterate through all the patients.
  - Test model performance on new patients

- For **imbalanced** dataset, do **not** just look at accuracy!
  - Example: for 1:9 positive/negative ratio, the predictor can achieve 0.9 accuracy if predicting **everything** as negative.
  - A good predictor should perform well for sensitivity, specificity and precision.
Test for Overfitting

- Example: selecting the number of nearest neighbors (K) for KNN

Smaller difference between training and testing errors $\rightarrow$ less overfitting
Benefit of Denoising Daily Features

- Singular Spectrum Analysis (SSA) denoises daily features for reliable predictions.

<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>0.7128 (0.0106)</td>
<td>0.6156 (0.0184)</td>
<td>0.3720 (0.0402)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>With SSA Denoising</td>
<td><strong>0.7402 (0.0075)</strong></td>
<td><strong>0.7283 (0.0220)</strong></td>
<td><strong>0.5360 (0.0265)</strong></td>
<td><strong>0.8611 (0.0000)</strong></td>
</tr>
</tbody>
</table>
Benefit of Fitbit + Clinical Features

- Machine learning models outperform traditional clinical scores (NSQIP).
- Wearable features complement clinical patient characteristics in prediction.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<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>0.7402 (0.0075)</td>
<td>0.7283 (0.0220)</td>
<td>0.5360 (0.0265)</td>
<td>0.8611 (0.0000)</td>
</tr>
<tr>
<td>Clinical Characteristics + Wearable Data</td>
<td>0.8691 (0.0061)</td>
<td>0.8786 (0.0076)</td>
<td>0.8120 (0.0183)</td>
<td>0.8556 (0.0111)</td>
</tr>
</tbody>
</table>

**NSQIP**: traditional rule-based surgical outcome evaluation tool

**Patient Characteristics**: the features extracted from medical records

**Wearable Data**: the features extracted from wearable data
Both Fitbit and clinical features contribute to predictions.
Machine Learning with Small Data

- Statistical analysis to assess predictability

- Feature extraction
  - Extract predictive features from time series data
  - Integrate clinical data and wearable features

- Mitigate overfitting in models
  - Adopt simple models
  - Reduce the number of variables through feature selection

- Evaluation
  - Don’t just look at accuracy, especially for imbalanced dataset
  - Assess overfitting by comparing performance on training/testing data
Wearable Use Cases

- Predicting clinical deterioration
  - Readmission and mortality of congestive heart failure patients
  - Post-operative complications and readmission of surgical patients

- Measuring mobility
  - Timed Up and Go with smartwatches

- Measuring mental health
  - Ecological momentary assessment (EMA)
  - Machine learning model based on physiological signals
Timed Up and Go with Smartwatch

- **Watch app**
  - Remind participants to take the assessment
  - Automatically upload the data to the cloud for analysis
  - Analyze gait and motion features
  - Feedback to physicians and participants

- Example: assess physical health and fall risk during prehabilitation

[https://www.cse.wustl.edu/~lu/TUG.mp4](https://www.cse.wustl.edu/~lu/TUG.mp4)

Developed by Ruixuan Dai (CSE)
Wearable Use Cases

- Predicting clinical deterioration
  - Readmission and mortality of congestive heart failure patients
  - Post-operative complications and readmission of surgical patients

- Measuring mobility
  - Timed Up and Go with smartwatches

- Measuring mental health
  - Ecological momentary assessment (EMA)
  - Model stress based on physiological signals
Measure Stress with EMA and PPG Sensor

Feeling in last 1 hour?

Stressed level? (1-4)

Happy level? (1-4)

10:01
09:59:12
Time to Assessment
Please get ready!

Objective Stress vs. Subjective Stress

- **Objective stress**: biological reaction to a stressful exposure
  - Physiological: response to stressor

- **Subjective stress**: subjective feeling of “being stressed”
  - Psychological: mental feeling

![Image of Physiological Signals](image-url)

- Objective stressor labels
- Objective stress detection model
- Subjective stress detection model
- Self-report labels

Feeling in last 1 hour?
- HAPPY
- STRESSED
- TIRED
- NEUTRAL
Physiological Signals vs Stress

Stress is related to heart rate, heart rate variability, respiratory rate

- Heart rate variability: Irregularities between heartbeats

- Heart rate ↑
- Respiratory rate ↑
- Heart rate variability ↓

Extracted features from PPG signals from a smart watch

<table>
<thead>
<tr>
<th>Category</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRV features</td>
<td>SDNN, RMSSD, SDSD, pNN20, pNN50, low frequency (LF) energy (0.04-0.15 Hz), high frequency (HF) energy (0.15-0.40 Hz), LF/HF energy ratio (LF_HF)</td>
</tr>
<tr>
<td>Non-HRV Inter-beat Interval features</td>
<td>mean, median, minimum, maximum, interquartile range (iqr), 20th percentile, 80th percentile, detrended fluctuation analysis (DFA), heart rate</td>
</tr>
<tr>
<td>Respiration features</td>
<td>RespWatch_RIIV [IoTDI 2021]</td>
</tr>
</tbody>
</table>
User Study for Stress Detection

- 32 participants subject to three types of stressors
  - Social stressor (public speech)
  - Cognitive stressor (arithmetic test)
  - Physical stressor (cold test)

- Respond to EMA about stress

Objective Stress

Subjective Stress

A. Resting video
   - Speech Prep.
   - Speech
   - Math1
   - Math2
   - Cold1
   - Cold2
   - Recover & EMA
   - Recover & EMA
   - EMA
   - Resting video

Duration (minutes):
- 20
- 4
- 4
- 5
- 4
- 5
- 1.5
- 1.5
- 20

B. Photoplethysmogram (PPG) at 200Hz
   - Inertial Measurement Units (IMU) at 50Hz
     - 3-axis accelerometer
     - 3-axis gyroscope

C. Pong game

Remaining Time: 3:25
Last number: 427
Correct: 3  Incorrect: 2

10/26/21
Personalized Subjective Stress Model

- **Subjective stress** is personal, with large variance among individuals

- Personalize model based on initial Perceived Stress Scale (PSS) score
  - Individual with a higher PSS score tend to report stress
  - Calibrate the ML probability output threshold based on the PSS score

---

**Training Set:** PPG features and stress labels

1. **Train generic subjective stress model**
2. **Find optimal personal threshold with highest F1-score**
3. **Ridge regression of optimal threshold and PSS score**
4. **Generic Subjective Stress Model + Threshold Regressor**

**New User**

**Pre-study PSS score from each user**
Evaluation

- Comparing objective stress model and subjective stress models
- Leave-one-subject-out (LOSO) cross-validation
  - SVM with rbf, Random Forest, AdaBoost, Logistic Regression
  - SVM achieves the best results

<table>
<thead>
<tr>
<th></th>
<th>ML Model</th>
<th>F-1 score</th>
<th>Accuracy</th>
<th>AUROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective stress</td>
<td>SVM with rbf</td>
<td>.623(.014)</td>
<td>.826(.073)</td>
<td>.790(.007)</td>
</tr>
<tr>
<td>Generic subjective stress</td>
<td>SVM with rbf</td>
<td>.520(.012)</td>
<td>.744(.102)</td>
<td>.719(.007)</td>
</tr>
<tr>
<td>Personalized subjective stress</td>
<td>SVM with rbf</td>
<td>.599(.014)</td>
<td>.810(.102)</td>
<td>.751(.011)</td>
</tr>
</tbody>
</table>

- Subjective stress models are less accurate than objective stress model.
- Personalized threshold improves the subjective stress detection.
Internet of Medical Things

- **Wearables**: continuous and nonobtrusive monitoring
- **Connectivity**: real-time monitoring and interactions
- **Cloud**: scalable to large cohorts
- **Analytics**: interpret data and predict outcomes

**Measuring** mobility and mental health

**Predicting** clinical outcomes

Just-in-time and adaptive **intervention**

**A powerful tool for precision medicine!**