Digital Phenotyping with Wearables

Learning from Wearable Data to Predict Clinical Outcomes

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!

Chenyang Lu
Wearables as Healthcare Tools

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

Two-way communication
ecological momentary assessments

Continuous, passive measurements
activity, heart rate, sleep, location…
Apple on Health

I believe, if you zoom out into the future, and you look back, and you ask the question, 'What was Apple's greatest contribution to mankind?', it will be about health.

--Tim Cook (Apple CEO)

Everyone says the biggest challenge in health is actually behaviour-change and awareness. They talk about the last mile and connecting with the patient. And when we've got hundreds of millions of phones in people's pockets and tens of millions of devices on people's wrists, plus trust from customers, well, this is an opportunity we can't squander.

--Jeff Williams (Apple COO)

HOW THE HEART BECAME THE CENTRE OF THE APPLE WATCH, Independent
https://www.independent.co.uk/life-style/gadgets-and-tech/features/apple-watch-health-heart-world-day-jeff-williams-interview-features-a9124601.html

10/27/2020
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 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)
  - Mental health model based on physiological signals
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
- Smart Phone
- Fitbit Cloud
- Transfer data
- Poll request
- HEROKU
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)

![Graph showing length of monitoring vs. number of participants]
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
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**
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

- **Challenges**
  - Imputation: Deal with missing data
  - Feature extraction: Extract predictive features from wearable data
  - Learning from small data: Develop generalizable models
  - Integration with clinical data
Feature Engineering Pipeline

- Avoid deep learning models due to small sample size
- Rely on feature engineering to extract predictive features as input to shallow machine learning models

Imputation with two-dimensional KNN

Daily feature extraction  High-level feature extraction

D: # of daily features
T: # of days

Inputs to predictive model

One-hot encoding

Demographics
Comorbidities

Heart rate  Step  Sleep stages

Missing value  Imputed value

HR_{t-5}  HR_{t-4}  HR_{t-3}  HR_{t-2}  HR_{t-1}  HR_{t}
Step_{t-5}  Step_{t-4}  Step_{t-3}  Step_{t-2}  Step_{t-1}  Step_{t}

Input
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
- 1st order: mean, max, min, skewness, kurtosis
- 2nd 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

**Important Features for Deterioration Early Warning**

10/27/2020
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><strong>HR LH</strong></td>
<td>4.0282</td>
<td>0.0566</td>
</tr>
<tr>
<td><strong>HR DFA 10</strong></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

- Impact of the number of nearest neighbors (K) on KNN performance

Smaller difference between training and testing errors $\rightarrow$ less overfitting
Result: Surgical Complications and Readmission

- Machine learning model outperformed clinical rules based on patient characteristics
- Integration of Fitbit features and clinical characteristics leads to significant performance improvement.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Precision</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSQIP with Clinical Characteristics</td>
<td>0.6078 (0.0000)</td>
<td>0.4008 (0.0000)</td>
<td>0.8400 (0.0000)</td>
<td>0.1389 (0.0000)</td>
<td>0.4038 (0.0000)</td>
<td>0.5455 (0.0000)</td>
</tr>
<tr>
<td>Clinical Characteristics</td>
<td>0.6932 (0.0196)</td>
<td>0.6441 (0.0252)</td>
<td>0.8440 (0.0120)</td>
<td>0.2297 (0.0676)</td>
<td>0.4265 (0.0231)</td>
<td>0.5663 (0.0212)</td>
</tr>
<tr>
<td>Wearable Data</td>
<td>0.7086 (0.0091)</td>
<td>0.6732 (0.0164)</td>
<td>0.8480 (0.0160)</td>
<td>0.3270 (0.0475)</td>
<td>0.4605 (0.0191)</td>
<td>0.5967 (0.0178)</td>
</tr>
<tr>
<td>Clinical Characteristics + Wearable Data</td>
<td><strong>0.8076 (0.0090)</strong></td>
<td><strong>0.7782 (0.0151)</strong></td>
<td><strong>0.8400 (0.0000)</strong></td>
<td><strong>0.5216 (0.0684)</strong></td>
<td><strong>0.5450 (0.0363)</strong></td>
<td><strong>0.6604 (0.0265)</strong></td>
</tr>
</tbody>
</table>

Machine learning model outperformed clinical rules based on patient characteristics.
Integration of Fitbit features and clinical characteristics leads to significant performance improvement.

Data collected from tele-monitoring devices has the power to predict surgical outcomes. The activity metrics are equally valuable to the medical record data. Moreover, the integration of these two contributes to the improvement of predictive performance.
Feature Importance

- Fitbit features and clinical characteristics are complementary in their predictive contributions.
Machine Learning with Small Data

- Apply 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 accuracy 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

- Assess physical health and fall risk during prehabilitation.
  - 20 participants undergoing neoadjuvant radiotherapy followed by surgery
  - Patients will complete TUG at home with the smartwatch for 90 days.

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

Joint work with Matthew Spraker (Radiation Oncology), 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

Joint work with Thomas Kannampallil (Anesthesiology, Informatics), Michael Avidan (Anesthesiology), Eric Lenze (Psychiatry), Ruixuan Dai (CSE)
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!