Predicting Post-Operative Complications with Wearables

Jingwen Zhang
02/08/2024
Pancreatic cancer is one of the most dangerous diseases:

• Low 5-year survival rate [American Cancer Society’s Cancer Facts and Figures 2022]

• The only cure is the surgery

“complications”: composite outcomes of readmission or severe complications

High rate of readmission and severe complications after the pancreatic surgery (40% - 60%) [Lubrano et al., 2018]
Surgical Outcome Prediction

**Goal:** reliable prediction of the post-operative complications

- **Decision support:** suitability for surgery
- **Intervention:** pre-habilitation
ACS-NSQIP

- Rule-based model used in practice leverages static clinical features (ACS-NSQIP)
  - Age, diabetes, hypertension...
- Predictive performance is moderate
Static Clinical Features

• Demographics
  • e.g. age, gender, race

• Comorbidity information
  • e.g. liver disease, chronic kidney disease

• Clinical presentations
  • e.g. Glucose, sepsis, albumin
Clinical Prediction with Wearables

Long-term, non-obtrusive monitoring

Static Clinical Features
From Medical Records

Wearable Data
(1 month before surgery)
From Fitbit

Time Series

Machine Learning

Post-operative Complications
(binary classification)
Web API Reference

Fitbit provides a set of public Web APIs that developers may use to retrieve Fitbit user data collected by the Fitbit trackers & smartwatches, Aria & Aria 2 scales, and manually entered log data. Anyone may use the Web APIs to build integrations with the Fitbit data services, so long as their application complies with the Fitbit Platform Terms of Service, the Fitbit User Data and Developer Policy, and the Fitbit user consents to share their data with the developer’s application.

Here are some resources to help you get started:

- **Developer Guide** contains information such as how to get started using the Web APIs, tips for designing your application, best practices and sample code.
- **Troubleshooting Guide** helps you debug common problems and errors, and tells you where to find some tools to help test and troubleshoot the endpoints.
- The API endpoint pages show you the proper syntax, the expected response and meaning, and any nuances that you need to be aware. The endpoints are grouped by the following:
  - **Active Zone Minutes Time Series** provides a user’s heart-pumping activity throughout the day.
  - **Activity and Activity Time Series** provides information about a user’s activity and their activity goals. The time series endpoints can be used to observe trends.
  - **Authorization** contains the endpoints needed to manage consent to Fitbit user data.
  - **Body and Body Time Series** provides data on a user’s weight and body fat percentage. The time series endpoints can be used to observe trends.
  - **Breathing Rate** provides a user’s average breaths per minute at night.
  - **Cardio Fitness Score (VO2 Max)** returns the maximum or optimum rate at which the user’s heart, lungs, and muscles can effectively use oxygen during exercise.
  - **Devices** contains information about the devices paired to a user’s account and when their last sync time.
  - **Electrocardiogram** contains information about the user’s on-device electrocardiogram readings.
  - **Friends** contains information about a user’s friends and their leaderboard.
  - **Heart Rate Time Series** returns a user’s heart rate and resting heart rate values. The time series endpoints can be used to observe trends.
  - **Heart Rate Variability** provides the room mean square of successive differences versus recorded during a user’s period of sleep.
  - **Intraday** provides a daily, granular level of detail to a user’s active zone minutes, activity, breathing rate, heart rate, heart rate variability and SpO2 metrics.
  - **Nutrition and Nutrition Time Series** allows a user to record the foods they consumed and include the food’s nutritional metadata. The time series endpoints can be used to observe trends.
  - **Sleep** returns information about a user’s sleep patterns.
  - **SpO2** provides a user’s blood oxygen levels.
  - **Subscription** notifies your application via webhooks when new data is available.
  - **Temperature** returns a user’s core and skin temperature.
  - **User** returns basic information about the user who shared their data with your application.
Challenges

• A small study cohort (61 patients)
• Fine-grained time series wearable data at minute granularity
Challenges

• A small study cohort (61 patients)
• Fine-grained time series wearable data at minute granularity
• Noisy and incomplete wearable data

Visualization of the missingness of heart rate for each patient
Previous Methods

- A small study cohort
- Fine-grained time series wearable data
- Noisy and incomplete wearable data

→ “Simple” machine learning models
→ Simple statistics of daily features [Bae et al., 2016, Low et al., 2019, Xu et al., 2019]
→ Imputation [Li et al., 2020]
  Simple daily features robust to missing data [Wang et al., 2018]
  Biorhythm features [Doryab et al., 2019]

- Lose temporal information
- May introduce extra bias
- Not good enough to describe the time series data
Feature Engineering Pipeline

I. Daily Feature Extraction

II. High-level Feature Extraction

- Step Count
- Heart Rate
- Sleep Stage

Wearable Time Series Data

D: Number of Daily Features
N: Number of Days

Inputs to predictive model

Static Clinical Features
Feature Engineering Pipeline

I. Daily Feature Extraction

- Step Count
- Heart Rate
- Sleep Stage

Wearable Time Series Data

D: Number of Daily Features
N: Number of Days

II. High-level Feature Extraction

Static Clinical Features

Inputs to predictive model
Missing Wearable Data

- **Step count**: always have values because the Fitbit device records steps as long as it is turned on
- **Heart rate**: may be complete, partially missing, or completely missing
  - Impute short missing segments
- **Sleep stage**: can be complete or completely missing for one night
  - No imputation for missing data
    - Not wearing the device
    - No heart rate
    - Zero step values
Extraction & Imputation Threshold

- **Extraction threshold**: Daily features are extracted when the daily yield $\geq$ extraction threshold (e.g. 8 hours)

- **Imputation threshold**: Impute short missing HR segments when the segment length $\leq$ imputation threshold (10 minutes) via k-nearest neighbors algorithm (KNN)
Time Series Data Imputation

- **Imputation threshold**: Impute short missing HR segments when the segment length $\leq$ imputation threshold with KNN

Hard to impute long missing segments with accurate values

Imputation for short missing segments of heart rate samples based on observed heart rate and steps
Feature Engineering Pipeline

I. Daily Feature Extraction

D: Number of Daily Features

N: Number of Days

II. High-level Feature Extraction

Inputs to predictive model

Wearable Time Series Data

Step Count

Heart Rate

Sleep Stage

D: Number of Daily Features

N: Number of Days

Static Clinical Features
I. Daily Feature Extraction

- **Semantic features**
  - Active/sedentary information, e.g., sedentary bout counts
  - Sleep status information, e.g., time of sleep,

- **Statistical features**
  - First-order features, e.g., skewness
  - Second-order features, e.g., inertia

- **Detrended fluctuation analysis features**
  - Capture local fluctuation

---

**Different missing rate?**

- **Daily active time**
- **Duration of wearing the device**
- **Daily active time**
  - # of complete data samples

**Standardization** is used to accommodate missing data
Feature Engineering Pipeline

I. Daily Feature Extraction

- Step Count
- Heart Rate
- Sleep Stage
- Wearable Time Series Data

D: Number of Daily Features
N: Number of Days

II. High-level Feature Extraction

Static Clinical Features

Inputs to predictive model
II. Denoise with SSA

Singular spectrum analysis (SSA)

• **Denoises** time series of daily features
II. Denoise with SSA

Singular spectrum analysis (SSA)

- **Denoises** time series of daily features
- Is **robust** to missing data

![Graphs showing trend extraction with complete and missing data samples](image)

- mean
- standard deviation
- slope
Clinical Study

Clinical study: 61 patients undergoing pancreatic surgery

<table>
<thead>
<tr>
<th>Label</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>“1” (Complications)</td>
<td>25</td>
</tr>
<tr>
<td>“0” (Without Complications)</td>
<td>36</td>
</tr>
</tbody>
</table>

Static Clinical Features

Wearable Time Series Data
**Different Data Source**

<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>

- Machine learning models outperform standard surgical risk scores
  - \(x^2\) AUPRC
  - \(x^3\) sensitivity at the same specificity
- Wearable data + clinical characteristics → best predictive performance

- **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
Patients without complications:

- The heart rate can rebound more rapidly following a significant change
- Sleep status is more stable
- More active
Influence of Extraction Threshold

<table>
<thead>
<tr>
<th>Extraction Threshold</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 h</td>
<td>0.8602 (0.0069)</td>
<td>0.8705 (0.0058)</td>
<td>0.7480 (0.0183)</td>
<td>0.8583 (0.0150)</td>
</tr>
<tr>
<td>2 h</td>
<td>0.8542 (0.0066)</td>
<td>0.8655 (0.0075)</td>
<td>0.7400 (0.0200)</td>
<td>0.8611 (0.0124)</td>
</tr>
<tr>
<td>3 h</td>
<td>0.8589 (0.0079)</td>
<td>0.8696 (0.0078)</td>
<td>0.7480 (0.0183)</td>
<td>0.8444 (0.0136)</td>
</tr>
<tr>
<td>4 h</td>
<td>0.8513 (0.0080)</td>
<td>0.8545 (0.0097)</td>
<td>0.7560 (0.0332)</td>
<td>0.8556 (0.0111)</td>
</tr>
<tr>
<td>5 h</td>
<td>0.8478 (0.0072)</td>
<td>0.8612 (0.0086)</td>
<td>0.7480 (0.0256)</td>
<td>0.8611 (0.0000)</td>
</tr>
<tr>
<td>6 h</td>
<td>0.8592 (0.0056)</td>
<td>0.8612 (0.0037)</td>
<td>0.7520 (0.0160)</td>
<td>0.8556 (0.0111)</td>
</tr>
<tr>
<td>7 h</td>
<td>0.8693 (0.0069)</td>
<td>0.8741 (0.0076)</td>
<td>0.7720 (0.0183)</td>
<td>0.8528 (0.0127)</td>
</tr>
<tr>
<td>8 h</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>0.8583 (0.0083)</td>
</tr>
<tr>
<td>9 h</td>
<td><strong>0.8825 (0.0071)</strong></td>
<td><strong>0.8870 (0.0061)</strong></td>
<td><strong>0.8160 (0.0196)</strong></td>
<td>0.8556 (0.0111)</td>
</tr>
<tr>
<td>10 h</td>
<td><strong>0.8817 (0.0036)</strong></td>
<td><strong>0.8878 (0.0070)</strong></td>
<td><strong>0.8360 (0.0120)</strong></td>
<td>0.8611 (0.0124)</td>
</tr>
</tbody>
</table>

Predictive Performance with Varying Extraction Threshold

Similar results within a range
Influence of Imputation Threshold

<table>
<thead>
<tr>
<th>Imputation Threshold</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (No imputation)</td>
<td>0.7277 (0.0089)</td>
<td>0.6953 (0.0194)</td>
<td>0.4720 (0.0466)</td>
<td>0.8556 (0.0111)</td>
</tr>
<tr>
<td>1 minute</td>
<td>0.7198 (0.0061)</td>
<td>0.6969 (0.0205)</td>
<td>0.5000 (0.0410)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>5 minutes</td>
<td>0.7294 (0.0087)</td>
<td>0.7082 (0.0177)</td>
<td>0.5200 (0.0400)</td>
<td>0.8556 (0.0111)</td>
</tr>
<tr>
<td>10 minutes</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>20 minutes</td>
<td>0.7337 (0.0080)</td>
<td>0.7169 (0.0226)</td>
<td>0.5440 (0.0445)</td>
<td>0.8528 (0.0127)</td>
</tr>
<tr>
<td>30 minutes</td>
<td>0.7313 (0.0119)</td>
<td>0.7033 (0.0248)</td>
<td>0.5440 (0.0543)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>60 minutes</td>
<td>0.7235 (0.0064)</td>
<td>0.6933 (0.0154)</td>
<td>0.4720 (0.0349)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>120 minutes</td>
<td>0.7392 (0.0108)</td>
<td>0.7204 (0.0239)</td>
<td>0.4880 (0.0431)</td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>240 minutes</td>
<td>0.7326 (0.0078)</td>
<td>0.7179 (0.0132)</td>
<td>0.4600 (0.0268)</td>
<td>0.8528 (0.0127)</td>
</tr>
</tbody>
</table>

Predictive Performance with Varying Imputation Threshold

Imputing long missing segments may introduce extra bias
### Evaluation: Pre-operative Duration

<table>
<thead>
<tr>
<th>Pre-operative Duration</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Dataset</td>
<td>0.8802  (0.0050)</td>
<td><strong>0.8871 (0.0087)</strong></td>
<td><strong>0.8320 (0.0160)</strong></td>
<td>0.8583 (0.0083)</td>
</tr>
<tr>
<td>5 Days for All the Patients</td>
<td><strong>0.8916 (0.0066)</strong></td>
<td><strong>0.8866 (0.0087)</strong></td>
<td>0.7760 (0.0265)</td>
<td>0.8583 (0.0083)</td>
</tr>
</tbody>
</table>

Predictive Performance with/without the Same Pre-operative Duration
Comparison to Baselines

<table>
<thead>
<tr>
<th>Method</th>
<th>AUROC</th>
<th>AUPRC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biorhythm Features [Doryab et al. 2019]</td>
<td>0.6267 (0.0000)</td>
<td>0.4835 (0.0000)</td>
<td>0.4000 (0.0000)</td>
<td>0.8611 (0.0000)</td>
</tr>
<tr>
<td>Min, Max, Mean of Daily Features [Li et al. 2020]</td>
<td>0.6844 (0.0101)</td>
<td>0.6211 (0.0318)</td>
<td>0.4520 (0.0360)</td>
<td>0.8528 (0.0127)</td>
</tr>
<tr>
<td>Our Feature Engineering Approach</td>
<td><strong>0.7326 (0.0074)</strong></td>
<td><strong>0.7192 (0.0154)</strong></td>
<td><strong>0.5480 (0.0440)</strong></td>
<td>0.8583 (0.0083)</td>
</tr>
</tbody>
</table>

Predictive Performance with Different Feature Extraction Methods on Wearable Data

**Better at handling missing and noisy time series data**

- Biorhythm features: cosinor analysis to describe the biorhythm
- Min, Max, Mean of daily features: forward imputation + statistics of daily features
Handling Missing and Noisy Data

• Impact of standardization when generating daily features (for raw time series data)

<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>

• Influence of SSA to denoise daily features (for 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
Conclusion

• Two-level feature engineering approach for wearable data
• Explore its feasibility in a clinical study with 61 patients (AUROC = 0.8802)
• Combine wearable data and clinical characteristics
• Extract expressive high-level features from noisy and incomplete wearable data
Reference

[1] American Cancer Society’s (ACS) publication, Cancer Facts & Figures 2022, the ACS website, the International Agency for Research on Cancer website, and the National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Program.


