Predicting Clinical Outcomes with Wearables

Machine Learning from Small Data

Chenyang Lu
Cyber-Physical Systems Laboratory
Department of Computer Science & Engineering
https://www.cse.wustl.edu/~lu/

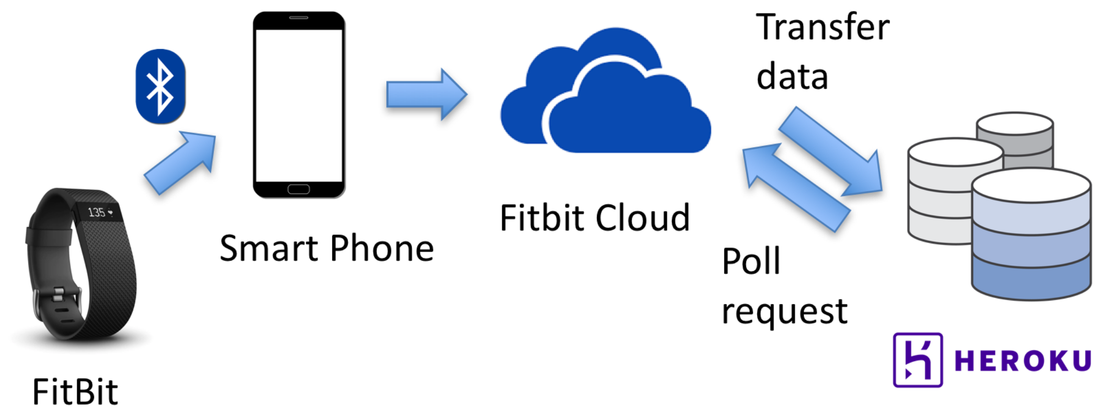
Machine Learning for Healthcare

- Machine learning: data-driven tools for healthcare.
  - Predicting clinical outcomes
  - Discover risk factors associated with outcome
  - Support clinical decisions to improve outcome

- Discover the insights behind diverse data
  - Mobile Health (mHealth)
    - Continuous monitoring outside hospitals
    - Small data: mHealth studies usually have moderate population
  - Electronic Health Record (EHR)
    - In-patient data
    - Big data: large patient population

Predict Readmissions with Fitbit

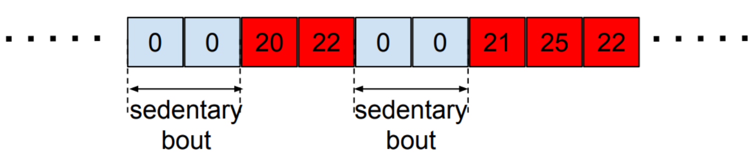
- Hospital readmission rate is high for heart failure patients.
  - ~25% patients readmitted within 30 days
- Predict deterioration (readmission+death) after discharge
  - Fitbit provides continuous monitoring of outpatients
  - Just-in-time intervention → better outcome and lower cost



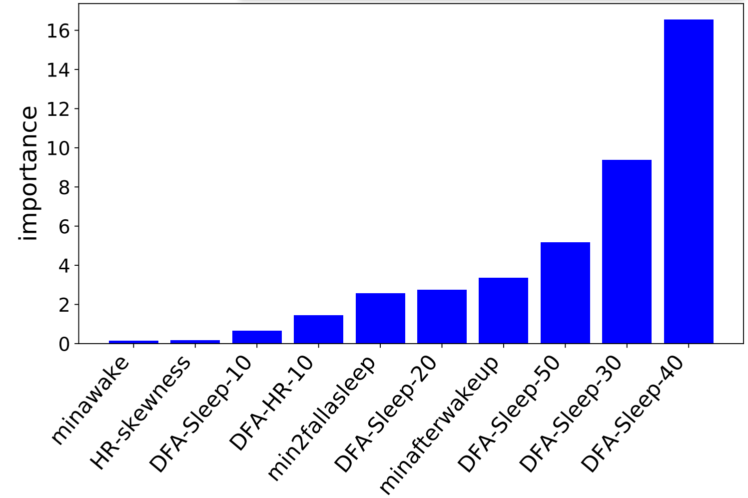
Joint work with Thomas Bailey (Infectious Diseases), Marin Kollef (Critical Care), Dingwen Li (CSE)
D. Li, J. Vaidya, M. Wang, B. Bush, C. Lu, M. Kollef and T. Bailey, Feasibility Study of Monitoring Deterioration of Outpatients Using Multi-modal Data Collected by Wearables, ACM Transactions on Computing for Healthcare, accepted.

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
  - 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

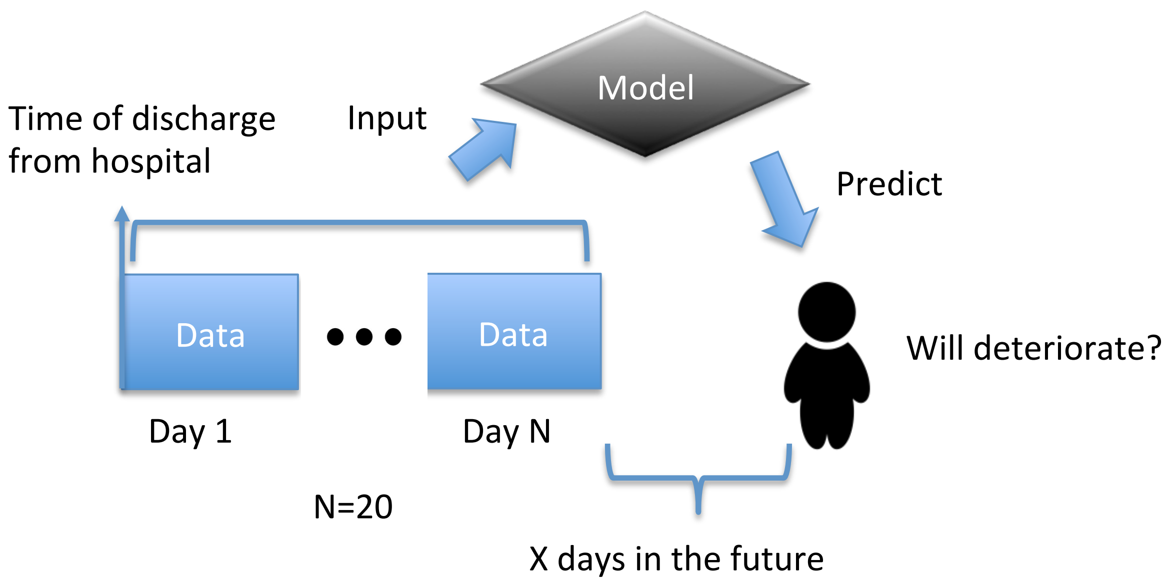


Important Features for Deterioration Early Warning



Predict Deterioration Risk

- **Input**: Fitbit data collected in the first N days since discharge
- **Deterioration**: composite outcome of either readmission or death
- **Small data**: 25 patients (18 with no deterioration vs. 7 deteriorated)



Assess Predictability

- Analysis of Variance (ANOVA) → predictability of outcomes
  - Test significant difference in the features between patients of different outcomes.
    - deteriorated patients vs. non-deteriorated patients
  - Assess feasibility of learning a stable predictor with the set of features.
- Significant differences in features → predictability

Features	F	p
HR skewness	7.9125	0.0099
HR correlation	5.4789	0.0283
HR DFA 10	5.3353	0.0302
Restless duration	5.2912	0.0308
Time in bed	5.2663	0.0312

Features with the largest F values.
High F-statistic and low p-value → significant difference between group means.

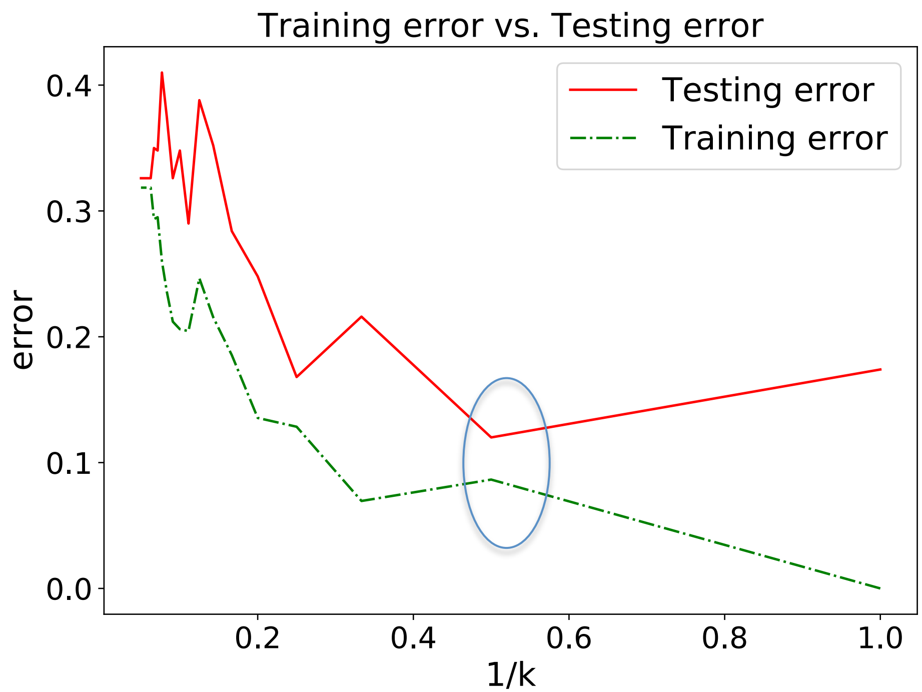
Feature Selection

- Select features using sequential forward feature selection
  - Avoid overfitting
  - Improve performance
- Features selected by the models have significant differences in ANOVA test

Features	F	p
sleep DFA 60	0.2254	0.6395
min asleep	**4.3128**	**0.0492**
daily step	**4.3625**	**0.0480**
restless count	**4.2324**	**0.0512**
awake count	**4.0429**	**0.0562**
min awake	2.2073	0.1509
HR LH	**4.0282**	**0.0566**
HR DFA 10	**5.3353**	**0.0302**

Test Overfitting

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



Smaller difference between training and testing errors → less overfitting

Evaluation with Small Data

- **Leave-one-out cross validation**
  - Leave a sample (or a group of samples) out for testing and train the model with rest of sample (or groups of samples).
  - Iterate through all the samples.

- 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.

$$Sensitivity = \frac{TP}{TP+FN} \quad Specificity = \frac{TN}{TN+FP} \quad Precision = \frac{TP}{TP+FP}$$

	Positive	Negative
Predicted positive	TP: true positives	FP: false positives
Predicted negative	FN: false negatives	TN: true negatives

Deterioration Risk Prediction

- Small data: 25 patients (18 with no deterioration vs. 7 with deterioration)
- KNN achieves higher specificity and precision than LACE index used in clinical practice.
- KNN performs better than neural network (NN)
  - **Simple** models help avoid overfitting on small data set.

Model	Sensitivity	Specificity	Precision	Accuracy
NN	0.0770	0.9459	0.3333	0.720
KNN	0.5385	**0.9820**	**0.9130**	**0.8667**
LACE	**0.7647**	0.6250	0.5556	0.7826

Machine Learning with Small Data

- Apply statistical analysis to assess outcome predictability with the features.

- Mitigate overfitting in models
  - Adopt simple models
  - Reduce the number of variables through feature selection
  - Use an ensemble of models (Gradient Boosted Trees, AdaBoost)

- Handle imbalanced data sets
  - Random oversampling of the minority class(es)
  - Random undersampling of the majority class(es)
  - Synthetic minority over-sampling technique (SMOTE)

- Evaluation
  - Don't just look at accuracy, especially for imbalanced dataset
  - Assess overfitting by comparing accuracy on training/testing data