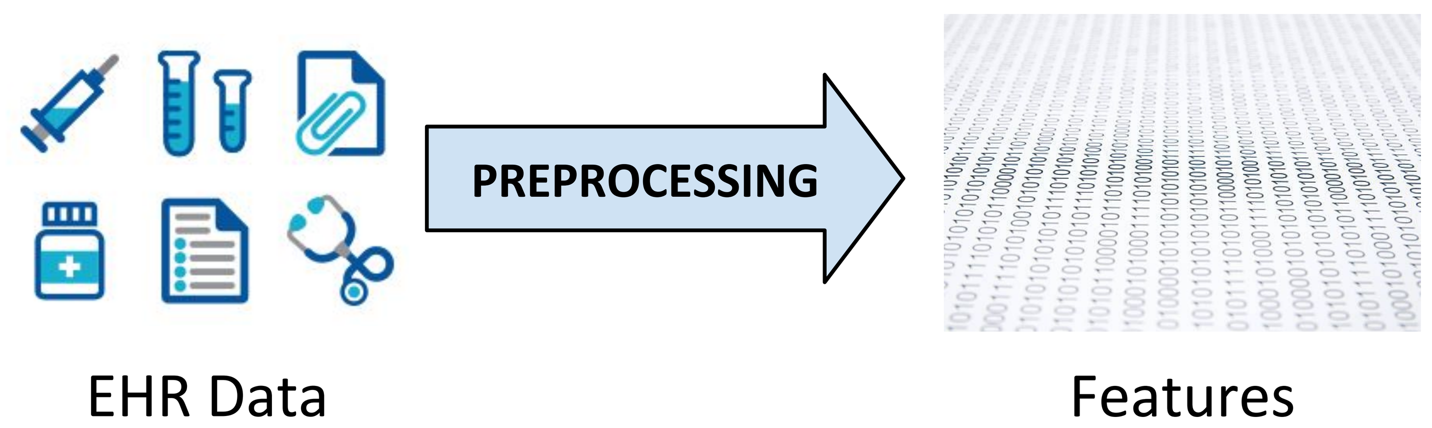
Democratizing EHR Analyses

A Comprehensive, Generalizable Pipeline for Learning from Clinical Data

Michael W. Sjoding MD MSc*, Shengpu Tang, Parmida Davarmanesh, Yanmeng Song, Danai Koutra PhD, Jenna Wiens PhD

UNIVERSITY OF MICHIGAN
CSE COMPUTER SCIENCE AND ENGINEERING
MICHIGAN MEDICINE
IHPI INSTITUTE FOR HEALTHCARE POLICY & INNOVATION

Motivation & Background



Challenges

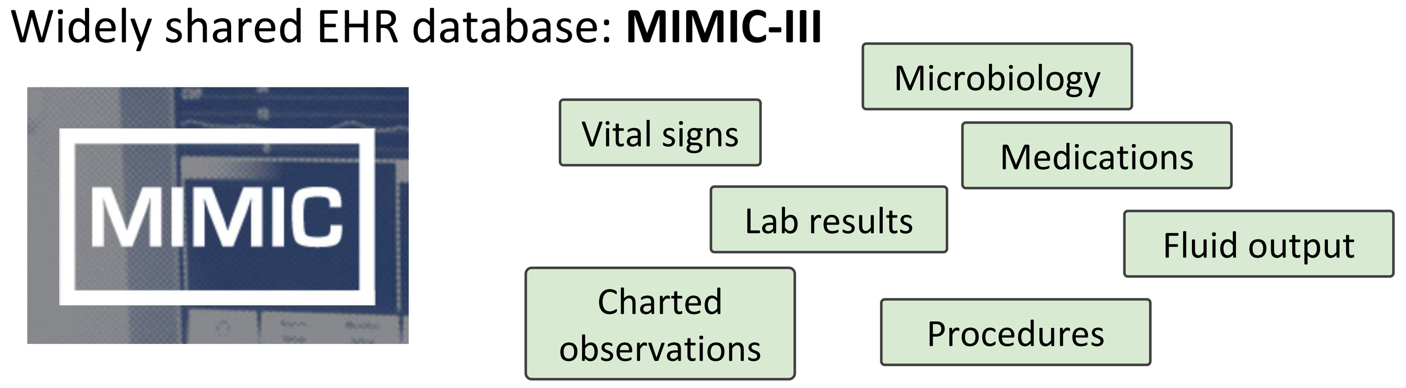
- **Messy**: high-dimensional, irregularly-sampled, multiple data types with frequent missingness
- **Many decisions** involved, labor-intensive, error-prone, and **ad-hoc**
- Heterogeneity makes **comparisons** difficult

Goals:

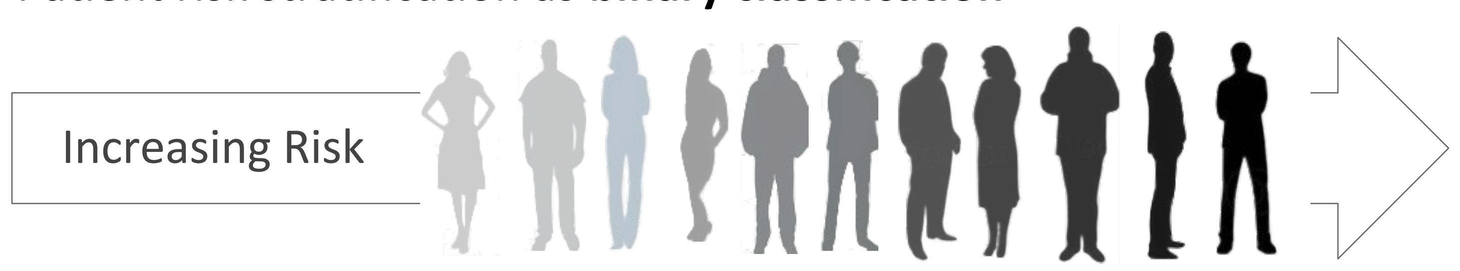
- **speed up** and **standardize** EHR data preprocessing
- an **open-source**, **generalizable**, data-driven pipeline
- offer a **quick** and **reasonable** starting point to build upon

Dataset & Tasks

Widely shared EHR database: **MIMIC-III**



Patient risk stratification as **binary classification**



Three **adverse outcomes**, five **prediction tasks**

- **In-hospital mortality**: using the first 48hrs of a patient's ICU visit (Harutyunyan et al., 2017)
- **Acute Respiratory Failure (ARF)**: need for respiratory support with positive pressure mechanical ventilation (Stefan et al., 2013)
- **Shock**: inadequate perfusion; receipt of vasopressor therapy (Avni et al., 2015)

Task	in-hospital mortality (48h)	ARF (4h)	ARF (12h)	shock (4h)	shock (12h)
N	8,577	15,873	14,174	19,342	17,588
%positive	12.0	18.3	9.7	14.9	7.8

Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. *Scientific Data* 2019;**6**(1):96.
Stefan MS, Shieh M-S, Pekow PS, et al. Epidemiology and outcomes of acute respiratory failure in the United States, 2001 to 2009: a national survey. *Journal of Hospital Medicine* 2013;**8**(2):76-82.
Avni T, Lador A, Lev S, Leibovici L, Paul M, Grossman A. Vasopressors for the treatment of septic shock: Systematic review and meta-analysis. *PLOS One* 2015;**10**(8):e0129305.
The Royal College of Physicians. *National Early Warning Score (NEWS): standardising the assessment of acute-illness severity in the NHS*. London: Report of a working party, 2012.

This work was supported by the Michigan Institute for Data Science (MIDAS); the National Science Foundation (NSF award no. IIS-1553146); and the National Heart, Lung, and Blood Institute (NHLBI grant no. R25HL147207).

FIDDLE - FlexIble Data-Driven pipeLinE

Input

Formatted data

ID	t	variable_name	variable_value
1	-	Age	59
1	-	Sex	M
1	0.2	Heart Rate	72
1	1.7	KCl_Amount	5.0
2	-	Age	45
2	1.2	Ciprofloxacin_Route	Oral
2	3.5	Temperature	38.5
3	3.6	WBC	6,000
3	2.5	Consciousness	Alert
4	8.9	Ventilator	Yes

...

User-defined arguments

Symbol	Description
T	time of prediction
dt	temporal granularity
θ_1	The threshold for ***Pre-filter***.
θ_2	The threshold for ***Post-filter***.
θ_{freq}	The threshold at which we deem a variable "frequent".
$\{\phi_j\}_{j=1}^{K}$	A set of K statistics functions (e.g., min, max, mean).

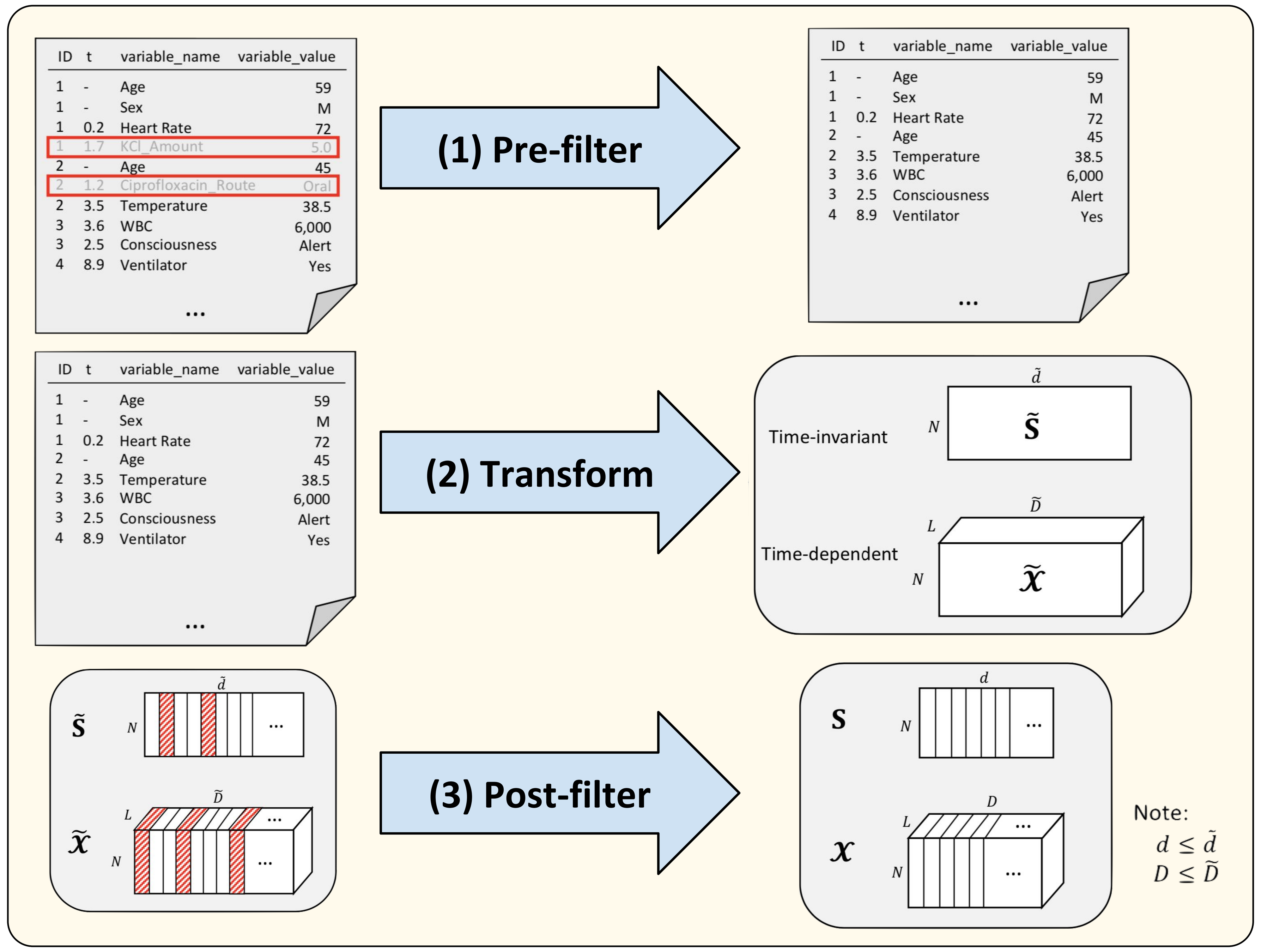
Other symbol definitions

Shapes & Sizes

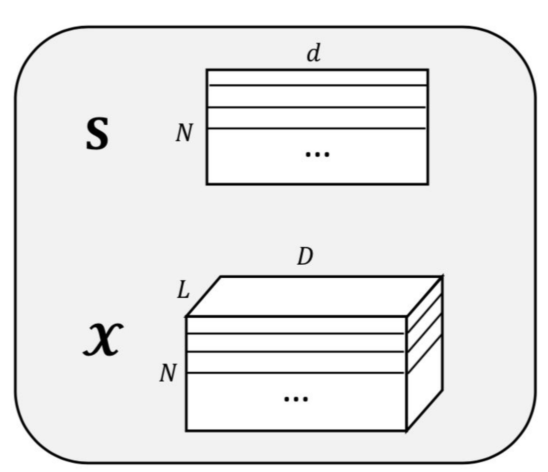
N number of examples
L temporal dimension / time bins
d number of time-invariant features
D number of time-dependent features

Data

$\mathbf{S}$ matrix of time-invariant features
$\mathcal{X}$ tensor of time-dependent features



Output



$$= \left\{ \begin{array}{l} s_i \in \mathbb{R}^{d} \\ x_i \in \mathbb{R}^{L \times d} \end{array} \quad \text{for } i = 1 \dots N \right\}$$

Experimental Results

Results: Extracted Features

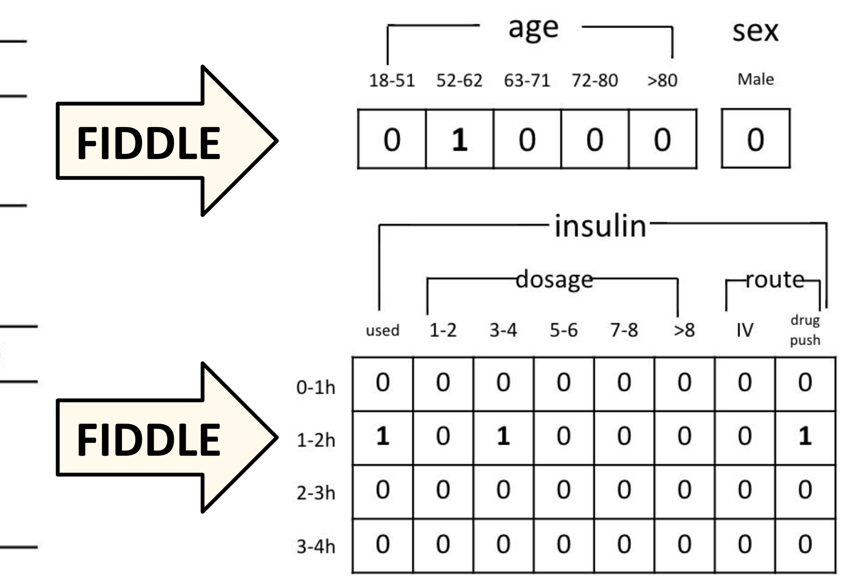
- Extracted 4,143-7,508 features on 8,577-17,588 ICU stays in 30-150 minutes

Time-invariant

ID	t	variable_name	variable_value
1	NULL	sex	female
1	NULL	age	55

Time-dependent

ID	t	variable_name	variable_value
2	1.5	insulin used	1
2	1.5	insulin dosage	3
2	1.5	insulin route	drug push



Experiments: Predictive Performance

Logistic Regression | Random Forest | 1D CNN | LSTM

- **In-hospital mortality:** an LSTM trained using the FIDDLE features outperformed the LSTM benchmark model (Harutyunyan et al., 2017)
- **ARF and Shock:** FIDDLE-LSTM outperformed the National Early Warning Score (NEWS) (The Royal College of Physicians, 2012)

AUROC Scores (95% confidence intervals)

Method	in-hospital mortality (48h) N=1,264	ARF (4h) N=1,823	ARF (12h) N=1,950	shock (4h) N=2,233	shock (12h) N=2,429
Baseline	0.839 (0.799, 0.877)	0.650 (0.614, 0.687)	0.628 (0.588, 0.666)	0.677 (0.644, 0.711)	0.682 (0.643, 0.721)
FIDDLE-LR	0.856 (0.821, 0.888)	0.733 (0.699, 0.767)	0.755 (0.717, 0.789)	0.775 (0.745, 0.805)	0.793 (0.758, 0.826)
FIDDLE-RF	0.814 (0.780, 0.847)	0.739 (0.703, 0.772)	0.759 (0.722, 0.793)	0.755 (0.725, 0.789)	0.773 (0.738, 0.807)
FIDDLE-CNN	0.886 (0.854, 0.916)	0.750 (0.718, 0.783)	0.768 (0.732, 0.801)	0.788 (0.761, 0.817)	0.795 (0.763, 0.826)
FIDDLE-LSTM	0.868 (0.835, 0.897)	0.744 (0.710, 0.777)	0.767 (0.732, 0.800)	0.777 (0.747, 0.808)	0.794 (0.761, 0.826)

FIDDLE -
an open-source pipeline for EHR preprocessing

tiny.cc/get_FIDDLE