

# MIMIC

## The Medical Information Mart for Intensive Care

MinDong Sung, M.D.

DHLab, Yonsei University College of Medicine

2021-07-01

# Introduction

# MIMIC-III

- Provide critical care data for over 40,000 patients admitted to ICU at the Beth Israel Deaconess Medical Center (BIDMC).
- deidentified, and patient identifiers were removed
- MIMIC-III has been integral in driving large amounts of research in clinical informatics, epidemiology, and machine learning.

# MIMIC-IV

- MIMIC-IV is intended to carry on the success of MIMIC-III
- Support a broad set of applications within healthcare.
- MIMIC-IV adopts a modular approach to data organization,
- Highlight data provenance
- Facilitate both individual and combined use of disparate data sources.

# Methods

## Acquisition

- Inclusion criteria
  - who were admitted to the BIDMC ED or any the ICU, 2008 - 2019.
- Extracted from the respective hospital databases
- A master patient list was created
- All source tables were filtered to only rows related to patients in the master patient list.

## Preparation

- The data were reorganized to better facilitate retrospective data analysis.
  - the de-normalization of tables
  - removal of audit trails
  - reorganization into fewer tables
  - to simplify retrospective analysis of the database.
- Data cleaning steps were not performed
  - to ensure the data reflects a real-world clinical dataset.

# Methods

## Deidentify

- Patient identifiers as stipulated by HIPAA were removed.
- Patient identifiers were replaced using a random cipher, resulting in deidentified integer identifiers for patients, hospitalizations, and ICU stays.
- Structured data were filtered using look up tables and allow lists.
- If necessary, a free-text deidentification algorithm was applied to remove PHI from free-text.
- Finally, date and times were shifted randomly into the future using an offset measured in days. A single date shift was assigned to each `subject_id`. As a result, the data for a single patient are internally consistent.

# Backgrounds for hospital data

How the patient data have been generated?

# CASE

Think about patient flows in hospital

- Patient
- Admission
- Chartevents
- Provider orders(poe)/medication administration(emar)
- Labevents
- Chest X ray
- Diagnoses\_icd
- Transfer
- chartevents
- datetimeevents
- inpuvents
- outpuvents
- procedureevents
- icustays



# Papers from MIMIC data

Researches based on MIMIC data

Clinical  
Model for clinicians  
Algorithm

# Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database (Annals of Intensive Care, 2018)

Vincent, J.L., Nielsen, N.D., Shapiro, N.I. et al. Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database. Ann. Intensive Care 8, 107 (2018).

# Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database (ICM, 2018)

Feng, M., McSparron, J.I., Kien, D.T. et al. Transthoracic echocardiography and mortality in sepsis: analysis of the MIMIC-III database. Intensive Care Med 44, 884–892 (2018).

# An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU (CCM 2018)

Nemati, Shamim et al., An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU, Critical Care Medicine: April 2018 - Volume 46 - Issue 4 - p 547-553

# Early prediction of circulatory failure in the intensive care unit using machine learning (Nature Medicine 2020)

Hyland, Stephanie L., et al. "Early prediction of circulatory failure in the intensive care unit using machine learning." *Nature medicine* 26.3 (2020): 364-373.

# Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques(Artificial Intelligence in Medicine, 2020)

Fen Miao et al., Continuous blood pressure measurement from one-channel electrocardiogram signal using deep-learning techniques, Artificial Intelligence in Medicine, Volume 108, 2020

# Clinical concept extraction using transformers (JAMIA 2020)

Xi Yang, Jiang Bian, William R Hogan, Yonghui Wu, Clinical concept extraction using transformers, Journal of the American Medical Informatics Association, Volume 27, Issue 12, December 2020, Pages 1935–1942,

# Tables



# Modular structure of MIMIC IV - Core

- **Patients**
  - Information that is consistent for the lifetime of a patient is stored in this table.
- **Admissions**
  - The admissions table gives information regarding a patient's admission to the hospital. Since each unique hospital visit for a patient is assigned a unique `hadm_id`, the admissions table can be considered as a definition table for `hadm_id`. Information available includes timing information for admission and discharge, demographic information, the source of the admission, and so on.
- **Transfers**
  - Physical locations for patients throughout their hospital stay.

# Modular structure of MIMIC IV - Hosp

- laboratory measurements (labevents, d\_labitems),
- microbiology cultures (microbiologyevents, d\_micro),
- provider orders (poe, poe\_detail),
- medication administration (emar, emar\_detail),
- medication prescription (prescriptions, pharmacy),
- hospital billing information (diagnoses\_icd, d\_icd\_diagnoses, procedures\_icd, d\_icd\_procedures, hcpcsevents, d\_hcpcs, drgcodes), and
- service related information (services).

# Modular structure of MIMIC IV

- icu
  - intravenous and fluid inputs (inputevents),
  - patient outputs (outputevents),
  - procedures (procedureevents),
  - information documented as a date or time (datetimeevents),
  - other charted information (chartevents).

# Other data related to MIMIC

# MIMIC-CXR (MIMIC IV)

1. Johnson AE, Pollard TJ, Berkowitz SJ, Greenbaum NR, Lungren MP, Deng CY, Mark RG, Horng S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. Scientific Data. 2019;6.
2. <https://github.com/MIT-LCP/mimic-cxr>

# MIMIC-Waveform (MIMIC III)

- Waveforms almost always include one or more ECG signals, and often include continuous arterial blood pressure (ABP) waveforms, fingertip photoplethysmogram (PPG) signals, and respiration, with additional waveforms (up to 8 simultaneously) as available.
- Numerics typically include heart and respiration rates, SpO2, and systolic, mean, and diastolic blood pressure, together with others as available.
- Recording lengths also vary; most are a few days in duration, but some are shorter and others are several weeks long.

<https://physionet.org/content/mimic3wdb/1.0/>

# How to access the data

# Reference

1. <https://physionet.org/content/mimiciv/1.0/>
2. <https://mimic.mit.edu/iv/>