의료와 데이터사이언스 8주차
(정형데이터 분석 3 – MIMIC)

서울대학교병원 헬스케어AI연구원
이현훈 교수

SNUH
서울대학교
SEOUL NATIONAL UNIVERSITY

CONTENTS

CONTENTS

Trends in Medical Data Science

SNUH

Open Dataset

- Publicly Available Datasets
- Open Source, Collaborations
- ICU Datasets
  - **MIMIC (MIT, US)**
  - K-MIMIC (SNUH, Korea)
  - eICU-CRD (MIT, US)
  - AmsterdamUMCdb (Netherland)
  - HiRID (Bern Univ, Switzerland)
  - SICdb (Salzburg Univ, Austria)
- Perioperative Datasets
  - VitalDB, INSPIRE (SNUH, Korea)
  - MOVER (UC Irvine, US)

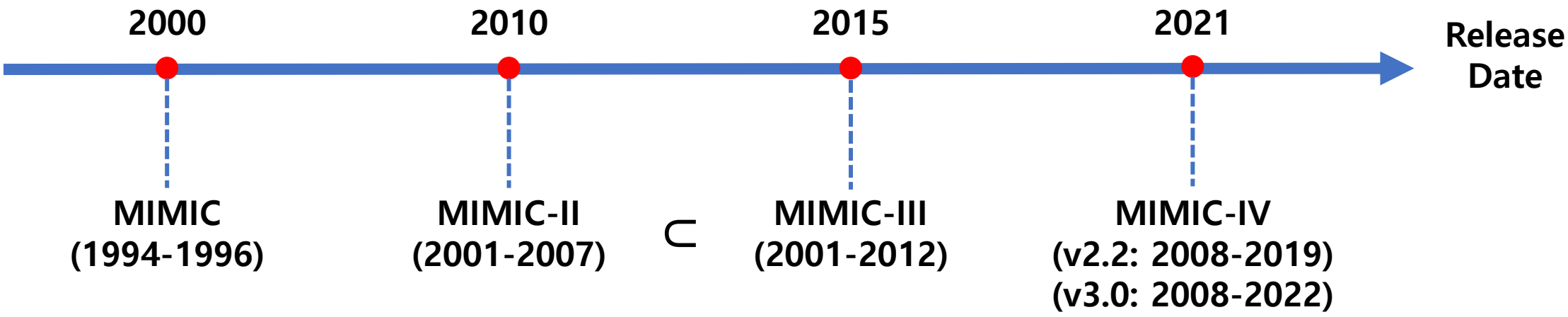
Federated Network

- Distributed, Restricted Data
- Moving Models and Queries
- Common Data Model: OMOP-CDM
- Research Networks
  - OHDSI Network
  - Research Border Free Zone
- Platforms
  - Mayo Clinic Platform (US)
  - FeederNet (47 hospitals, 57M pts, Korea)

H.C. Lee. (2023.10) Perioperative Data Science, Global trends and Future Directions

MIMIC Dataset

- **M**edical **I**nformation **M**art for **I**ntensive **C**are
- A large, single-center database comprising information relating to patients admitted to critical care units at a large tertiary care hospital



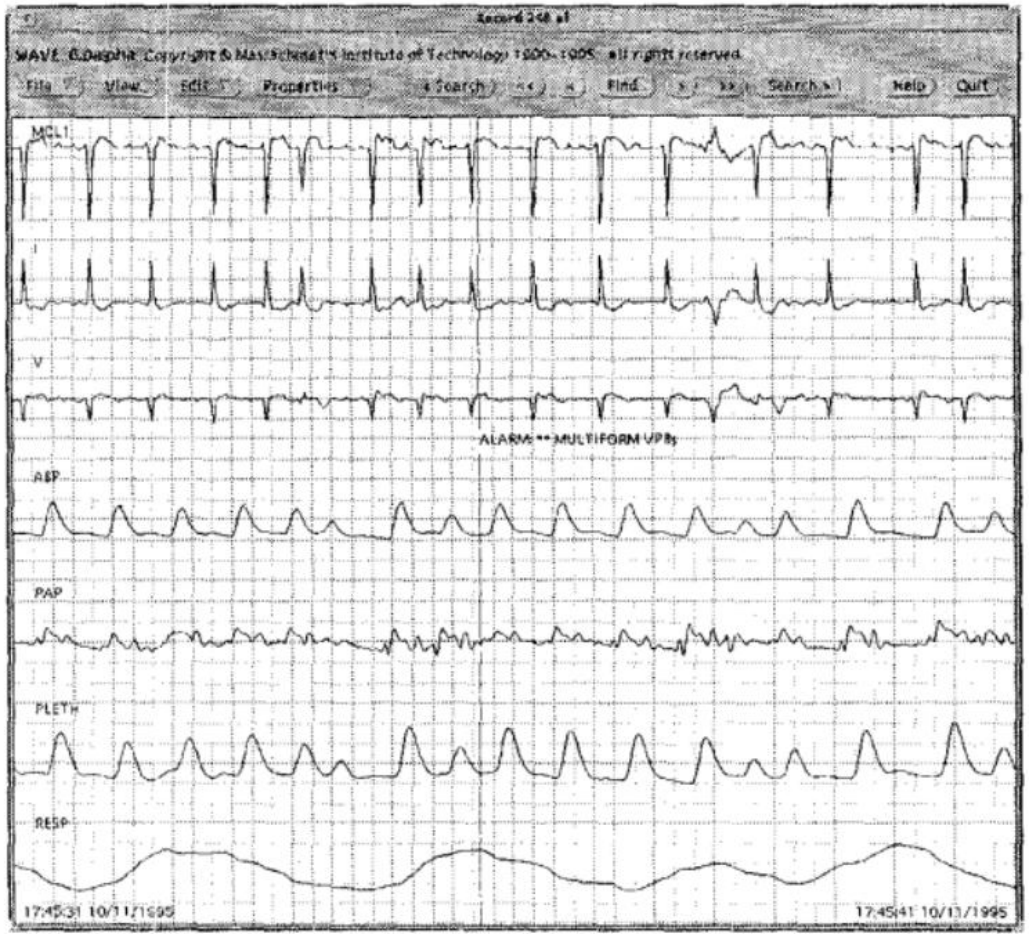
MIMIC I

Multiparameter Intelligent Monitoring for Intensive Care

- In 1996, 90 ICU patients
- 20 hours of ECG, ABP, PAP, PLETH
- First attempt to build a collection of multi-parameter recordings of ICU patients

Roger Mark

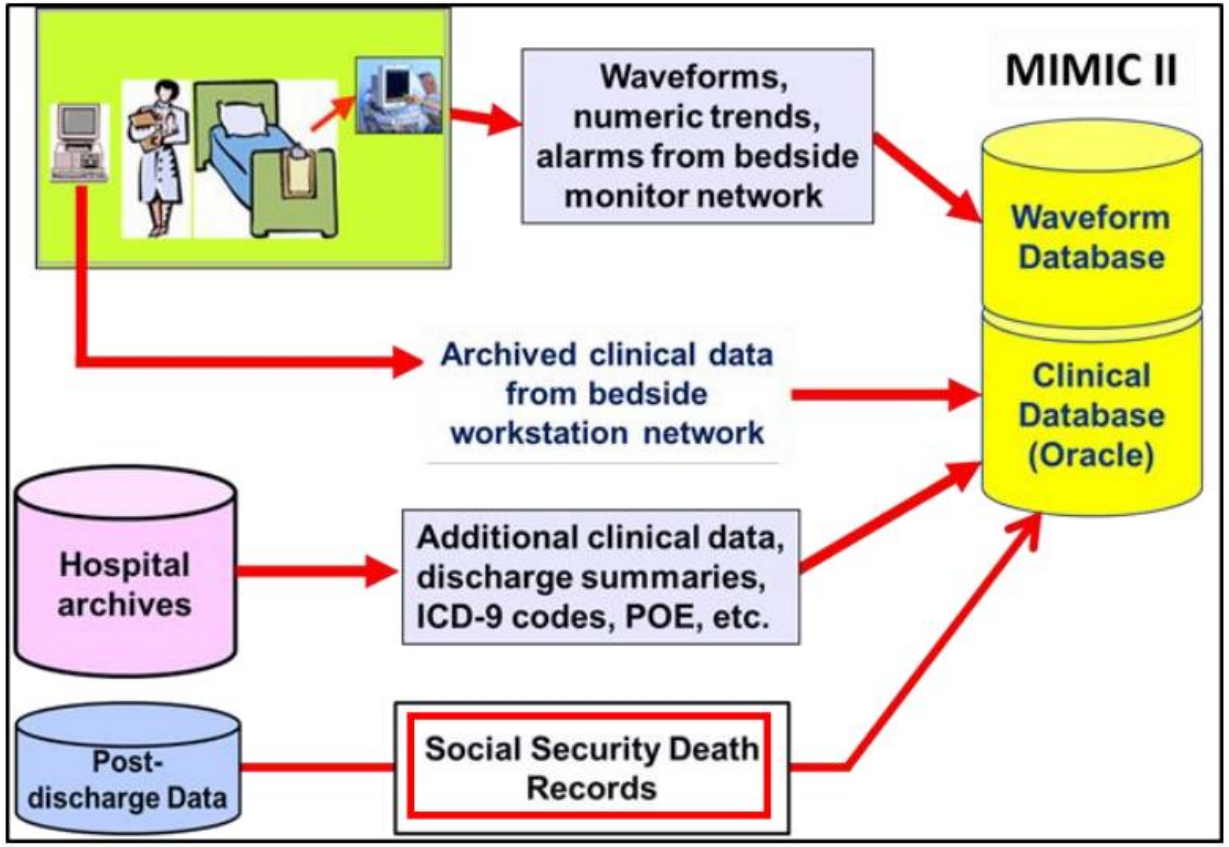
George Moody



Moody GB, Mark RG, A Database to Support Development and Evaluation of Intelligent Intensive Care Monitoring, Computers in Cardiology 1996

MIMIC II

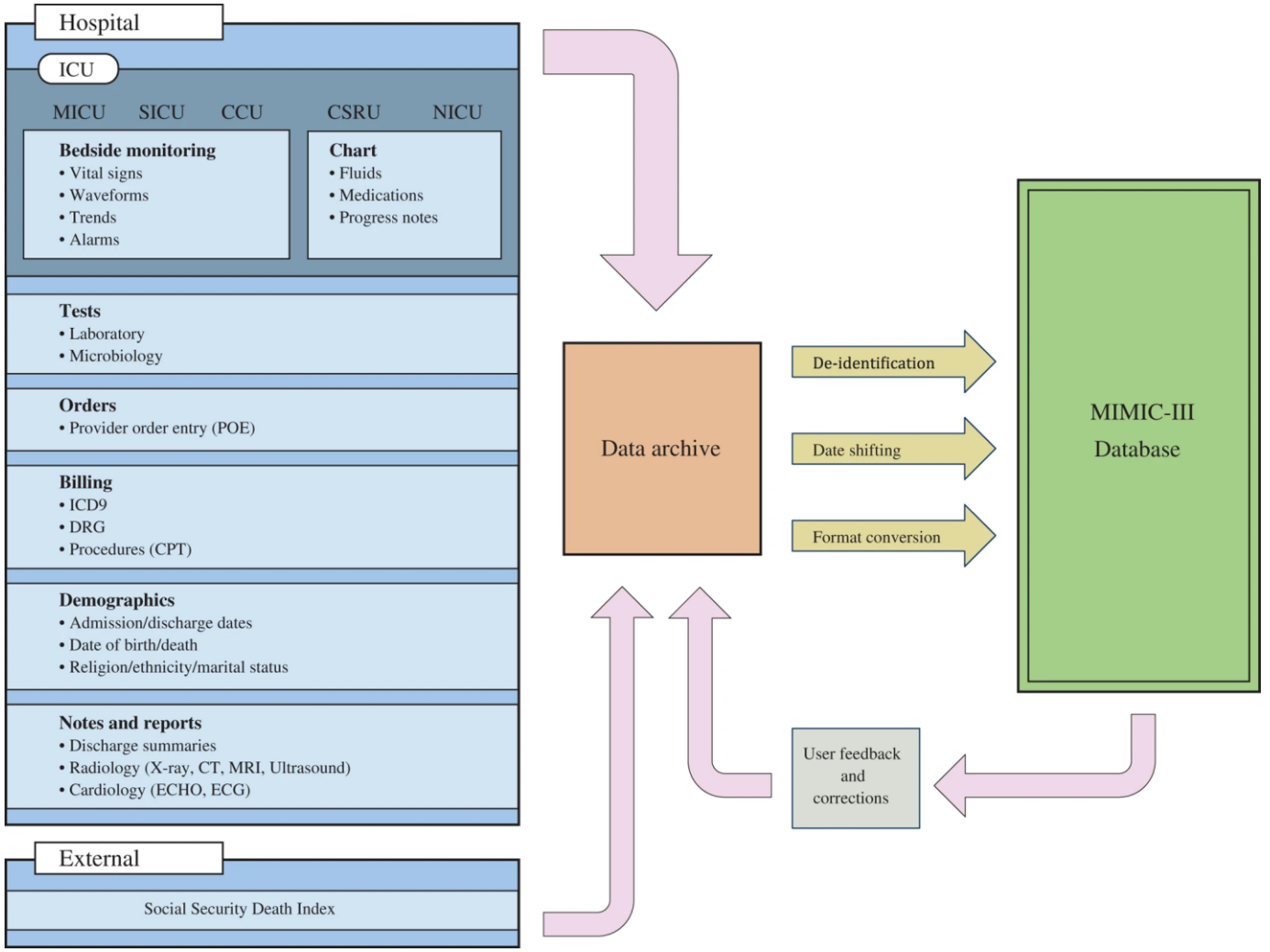
- In 2003, MIT, Philips, BIH received NIH fund
- Clinical data were collected between 2001 and 2007 from a variety of ICUs (MICU, SICU, CCU, NICU)
- Publicly released in 2010
- 25,328 ICU stays, 22,870 admissions



Joon Lee, et al. Open-access MIMIC-II database for intensive care research. Annu Int Conf IEEE Eng Med Biol Soc, 2011

MIMIC III

- In 2015, major update: includes **clinical notes**
- Renamed to the "Medical Information Mart for the Intensive Care"
- >40,000 patients, >50,000 ICU stays, between 2001-2012
- Cited by >**7000 publications (24.07.31)**

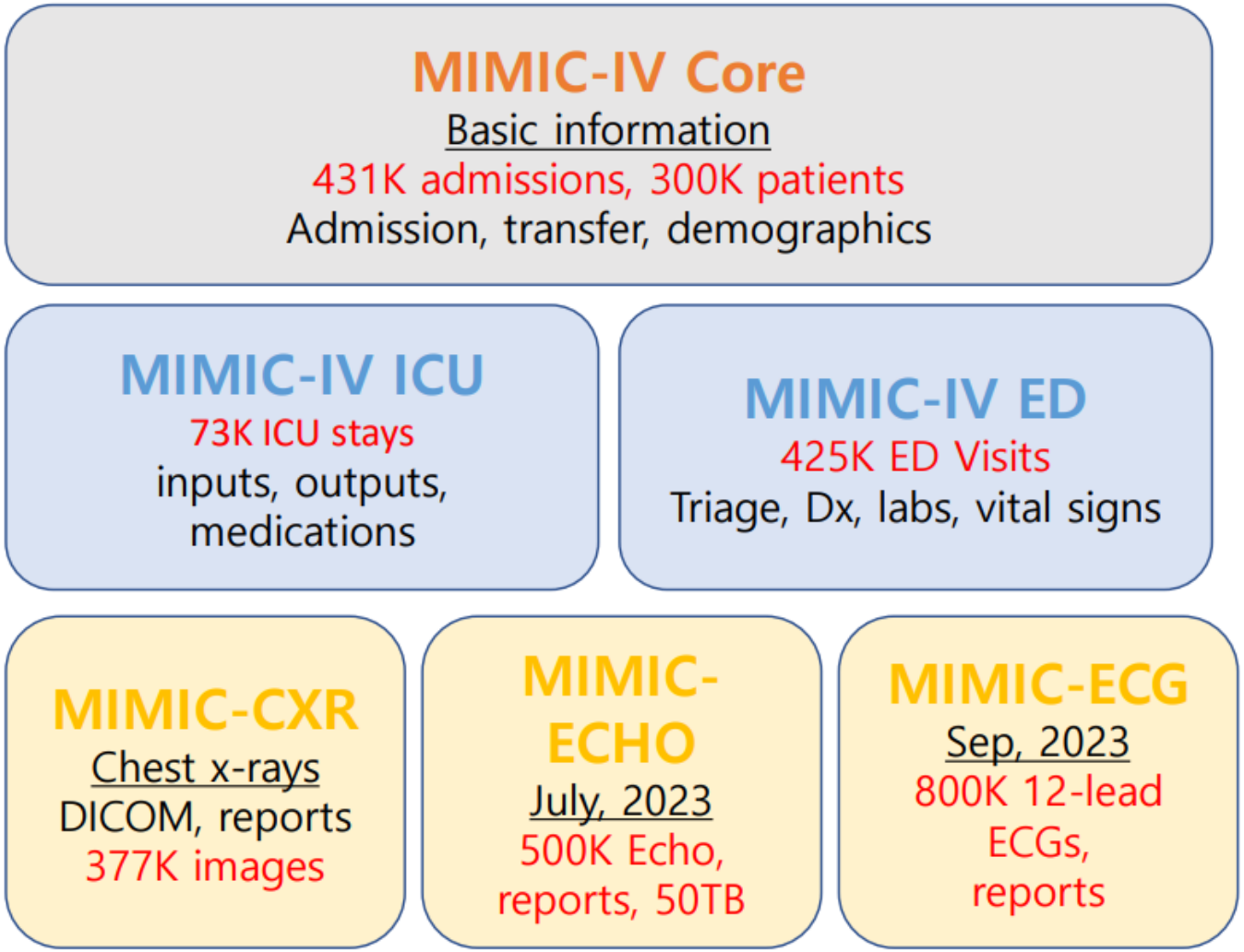


Johnson, A., Pollard, T., Shen, L. *et al.* MIMIC-III, a freely accessible critical care database. *Scientific Data*, 2016.

MIMIC IV

SNUH

- In March **2021**
- between 2008 – 2019 (overlapped patients btwn 2008-2012)
- Admissions: 431,231
  Patients: 299,712
  ICU stays: 73,181 (v2.2)
- **Modular approach**

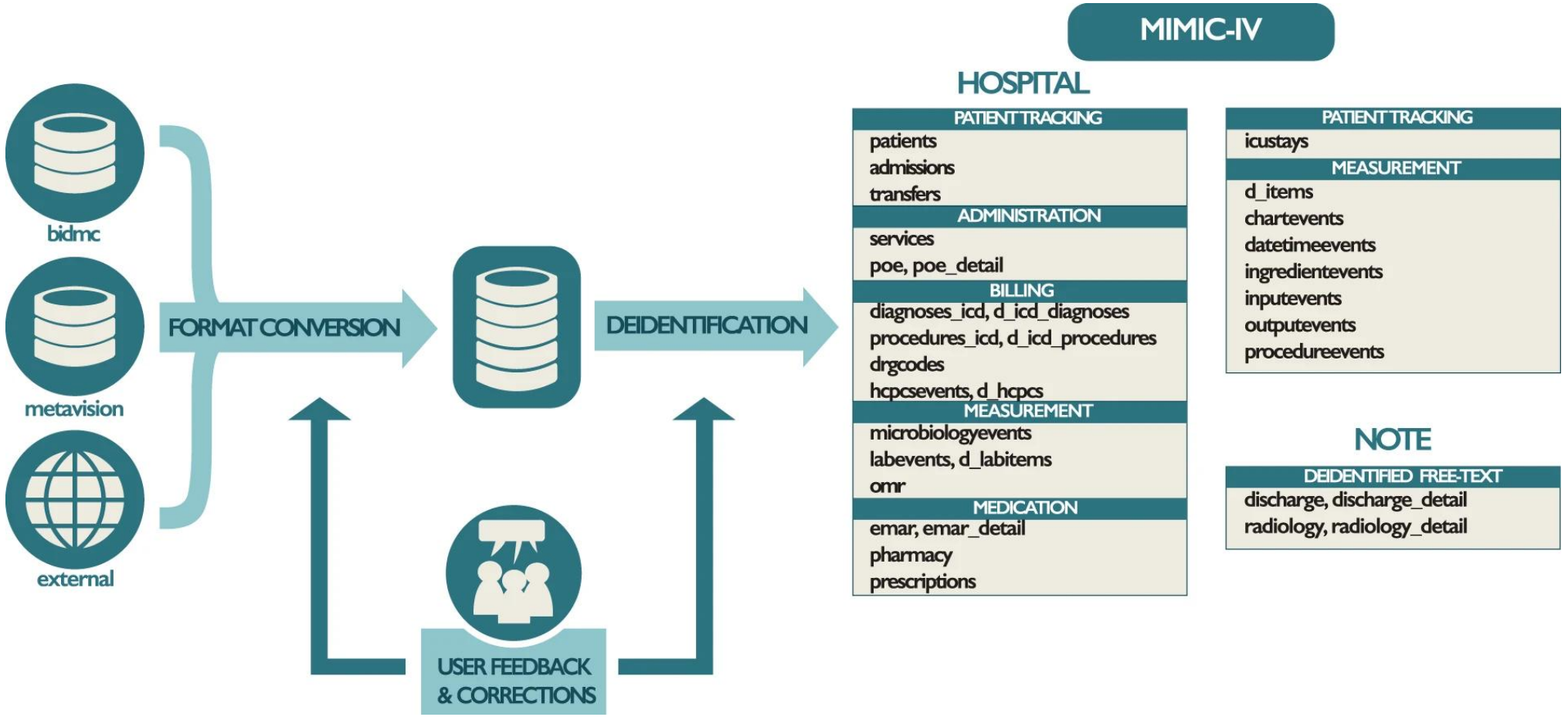


Alistair Johnson et al. MIMIC-IV, a freely accessible electronic health records dataset. Scientific Data 2023.

Patient composition

- **MIMIC-IV (3.0)** : patients admitted to the ED and the ICU
  - **364,627 unique patients**
    - **223,452 patients** : at least one hospitalization
    - **141,175 patients** : only seen in the ED

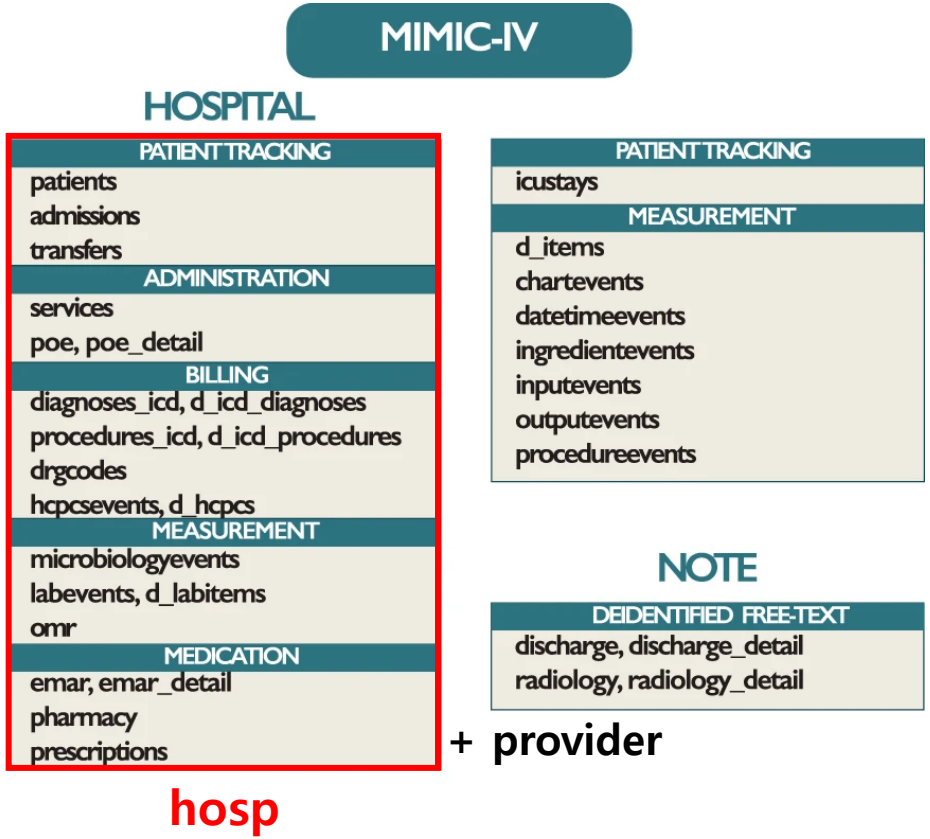
Dataset collection process



BIDMC: Beth Israel Deaconess Medical Center

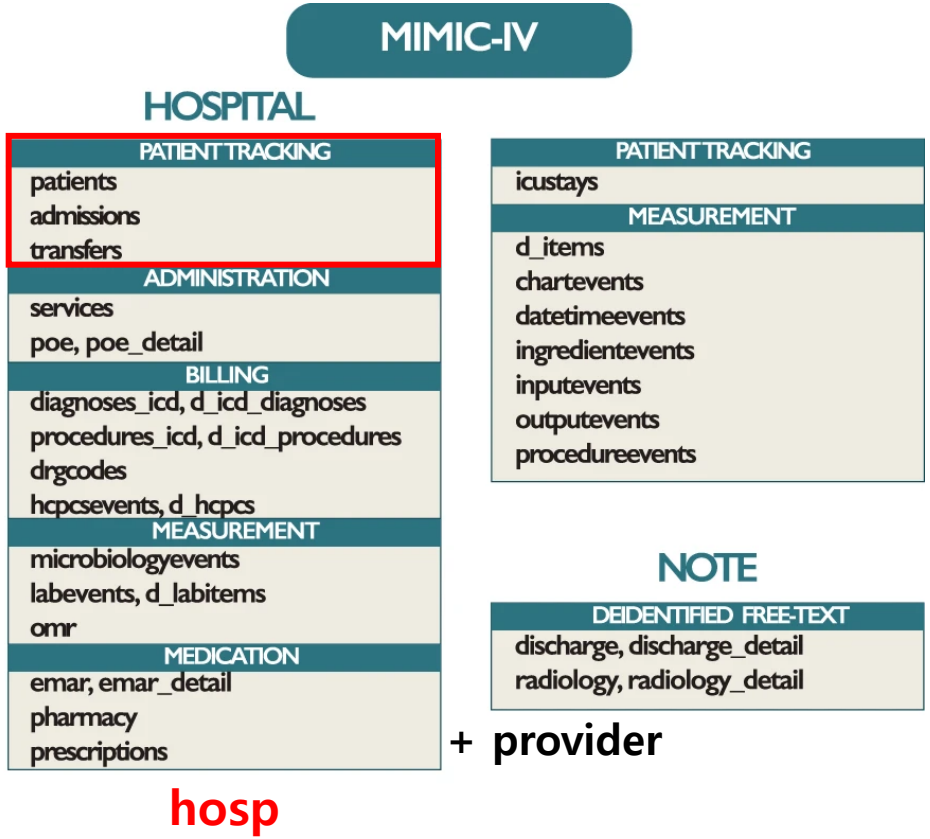
Alistair Johnson et al. MIMIC-IV, a freely accessible electronic health records dataset. Scientific Data 2023.

Data description - *hosp* module



- **546,028** unique hospitalizations, **223,452** unique patients
- **Sources**: BIDMC EHR
- **Identifiers**
  - subject_id
  - hadm_id : a single hospitalization
    - NaN value – outside of an inpatient encouter
  - item_ids : linkage to d_{table} (description)

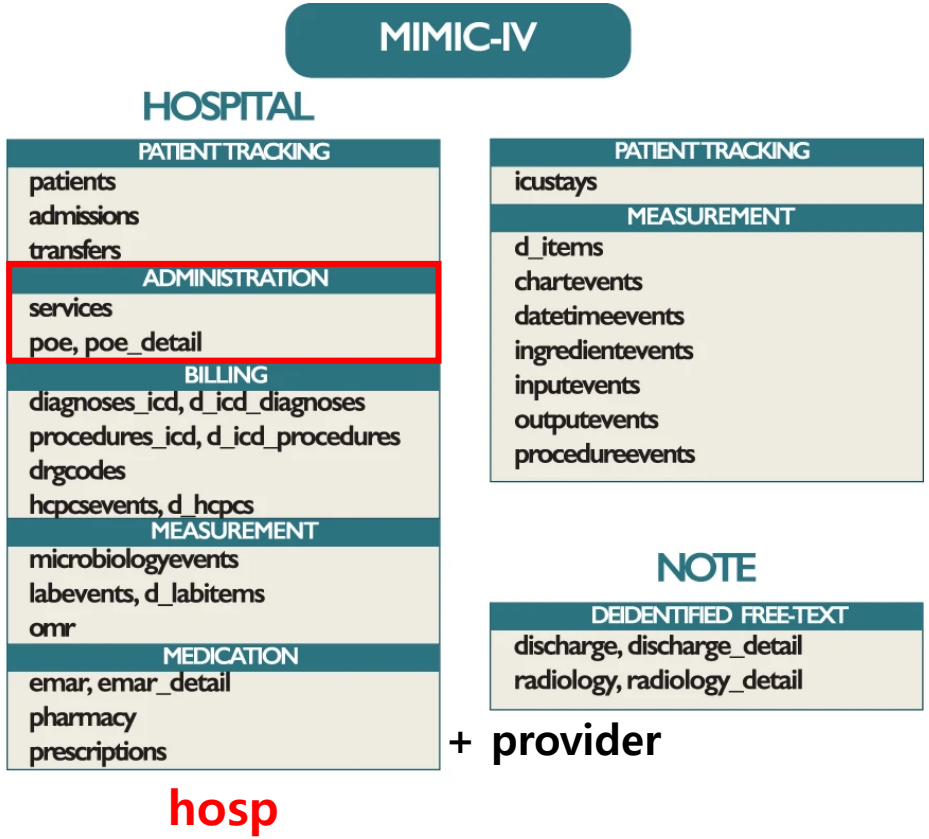
Data description - *hosp* module



- **Patient tracking**
  - ***patients*** : patient demographics
    - Patient's administrative gender, age, date of death
  - ***admissions*** : hospitalizations
  - ***transfers*** : intra-hospital transfers

	subject_id	hadm_id	transfer_id	eventtype	careunit	intime	outtime
2	10000032	22595853	33258284	ED	Emergency Department	2180-05-06 19:17	2180-05-06 23:30
3	10000032	22595853	35223874	admit	Transplant	2180-05-06 23:30	2180-05-07 17:21
4	10000032	22595853	36904543	discharge	UNKNOWN	2180-05-07 17:21	
5	10000032	22841357	34100253	discharge	UNKNOWN	2180-06-27 18:49	
6	10000032	22841357	34703856	admit	Transplant	2180-06-26 21:31	2180-06-27 18:49
7	10000032	22841357	38112554	ED	Emergency Department	2180-06-26 15:54	2180-06-26 21:31
8	10000032	25742920	35509340	admit	Transplant	2180-08-06 1:44	2180-08-07 17:50
9	10000032	25742920	35968195	ED	Emergency Department	2180-08-05 20:58	2180-08-06 1:44
10	10000032	25742920	38883756	discharge	UNKNOWN	2180-08-07 17:50	

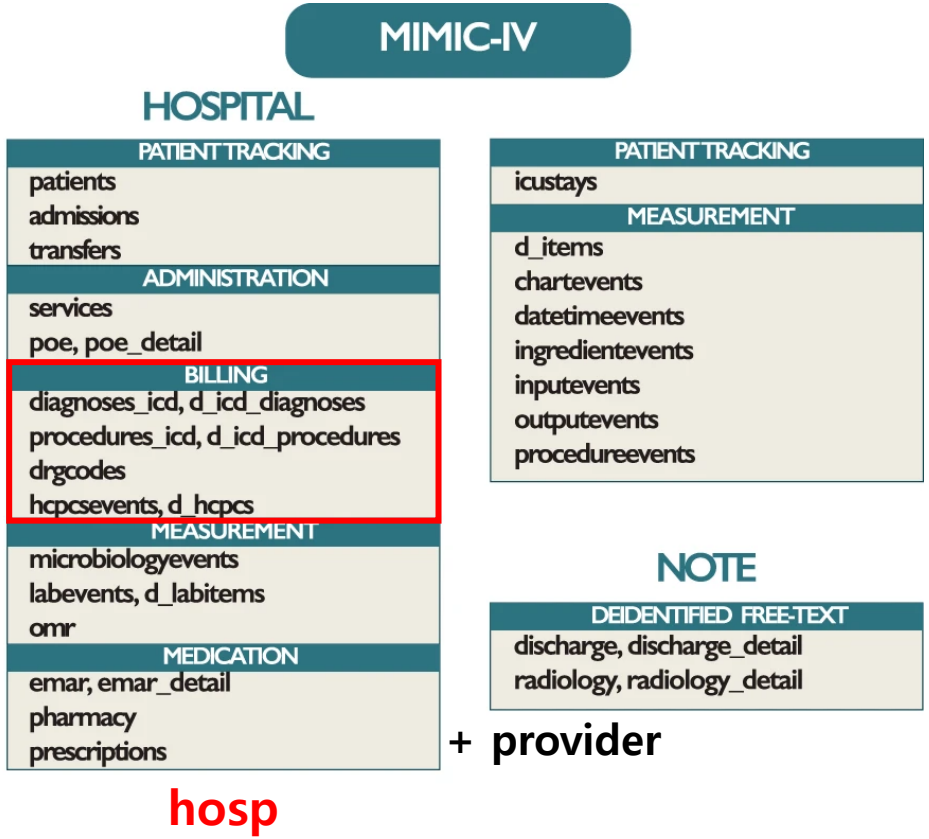
Data description - *hosp* module



- **Administration**
  - ***services*** : hospital-related services related information
  - ***poe***, ***poe_detail*** : orders made in the provider order entry (POE) system *
    - Provide the date and time of an order

* POE system : used within the hospital to make orders related to diagnoses, imaging, consultation, treatment

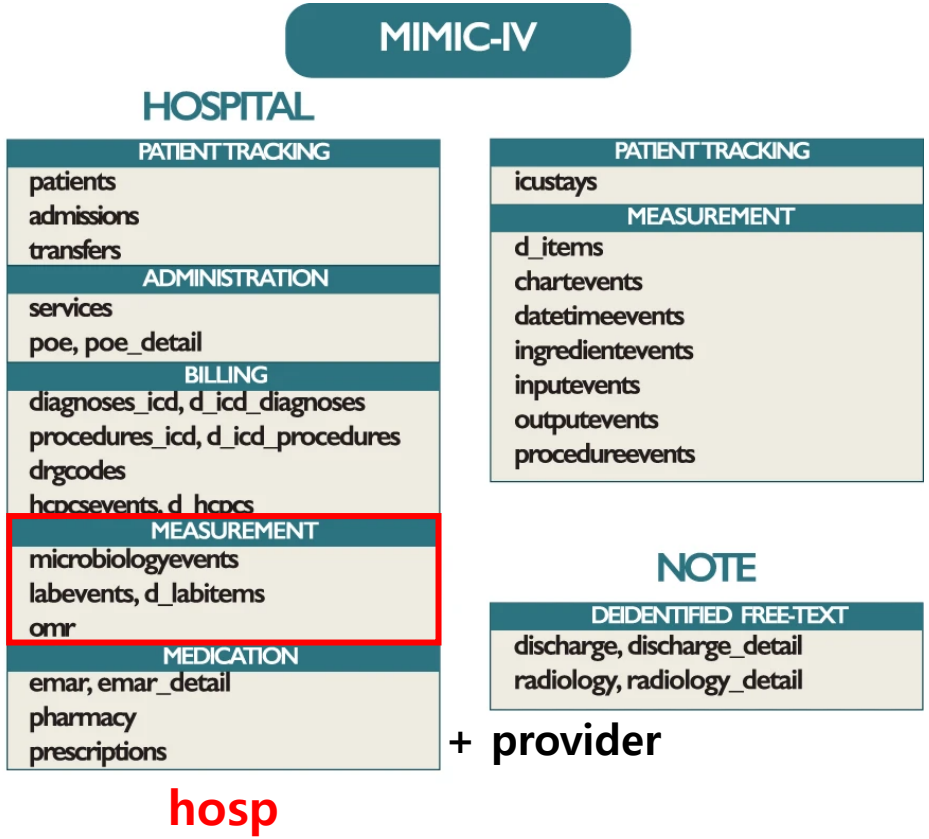
Data description - *hosp* module



- **Billing**
  - ***diagnoses_icd*** : coded diagnoses representing the hospitalization
    - Ontology: ICD-9-CM, ICD-10-CM
    - *d_icd_diagnoses* : definitions for ICD codes
  - ***procedures_icd*** *:* coded procedures
    - Ontology: ICD-9-PCS, ICD-10-PCS
  - ***drgcodes*** : Diagnosis Related Groups codes *
  - ***hcpcevents*** : billing by the hospital for provided services (ex. mechanical ventilation)

* DRG: billable codes used to assign an overall cost to a hospitalization

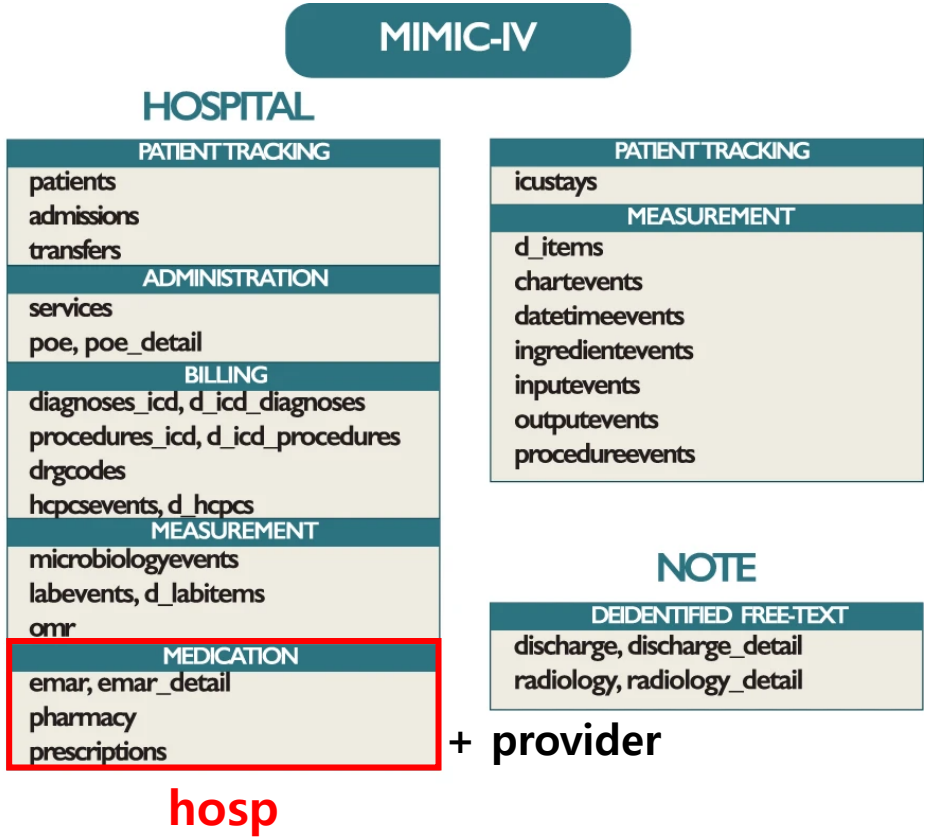
Data description - *hosp* module



- **Measurement**
  - ***microbiologyevents*** : microbiology measurements
  - ***labevents*** *:* laboratory measurements
    - *d_labitems*: definitions for concepts in *labevents*
  - ***omr*** : information from the Online Medical Record (OMR) *
    - Five measurements: blood pressure, height, weight, body mass index, eGFR
    - Both inpatient, outpatient visits
      - Including 'baseline value' before hospitalization

* OMR: a general system used for documenting patient information from visits at BIDMC affiliated institutes

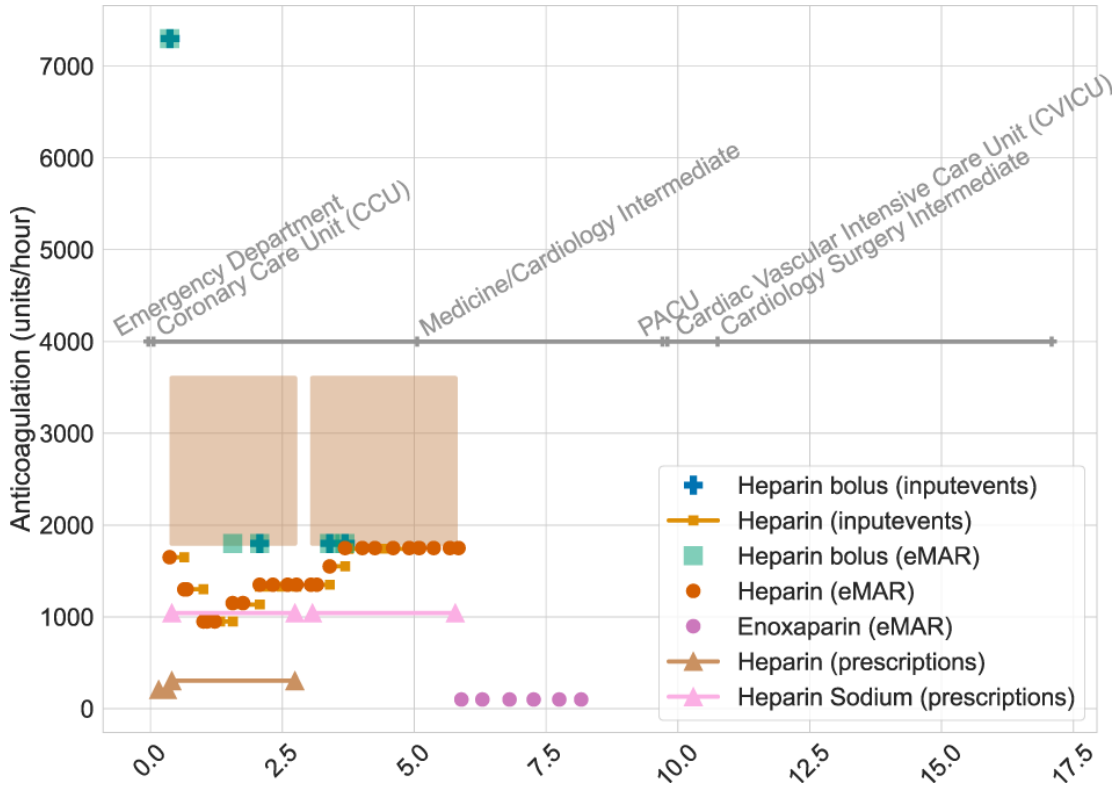
Data description - *hosp* module

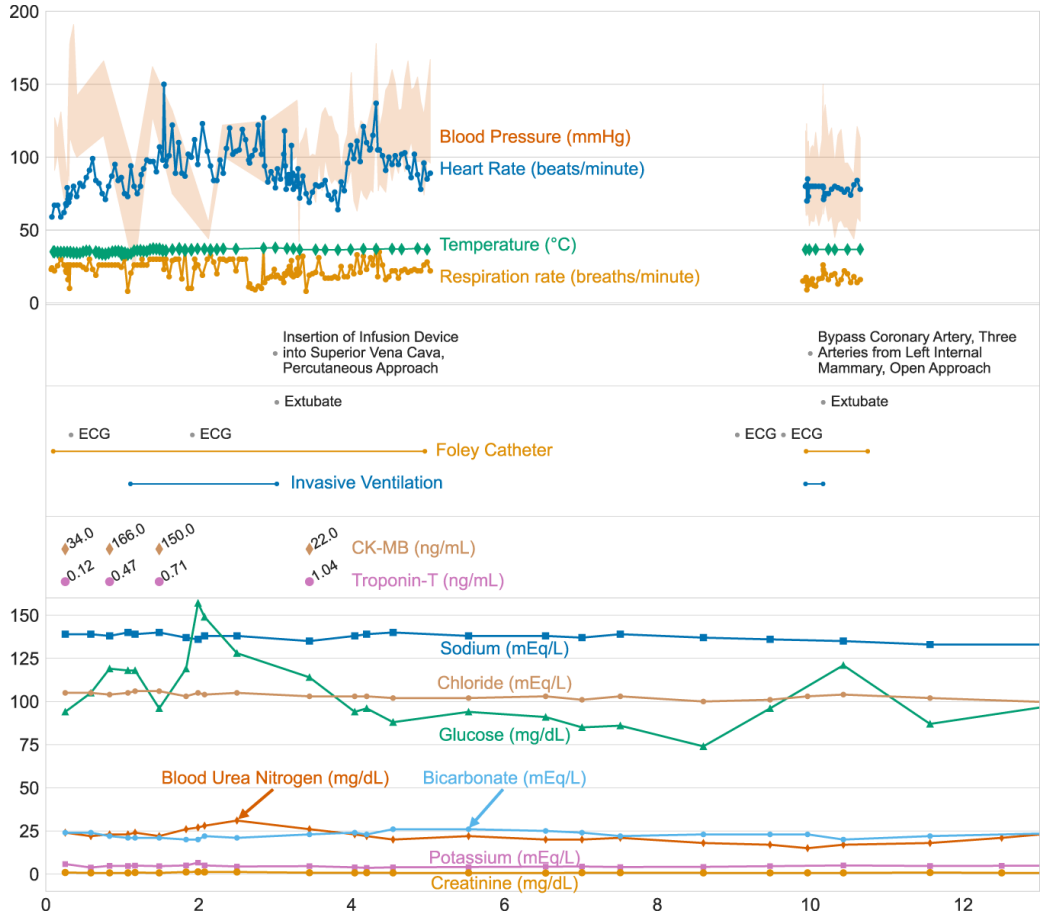


- **Medication**
  - ***prescriptions*** : **'order'** made by a provider
  - ***pharmacy*** *:* detailed information regarding the compoud prescribed
  - ***emar*** : **'administration'** records from the electronic Medicine Administration Record (eMAR)
    - 2014-2016 : first deployed => By 2016, all units
    - Link with *poe* (poe_id), *pharmacy* (pharmacy_id) table

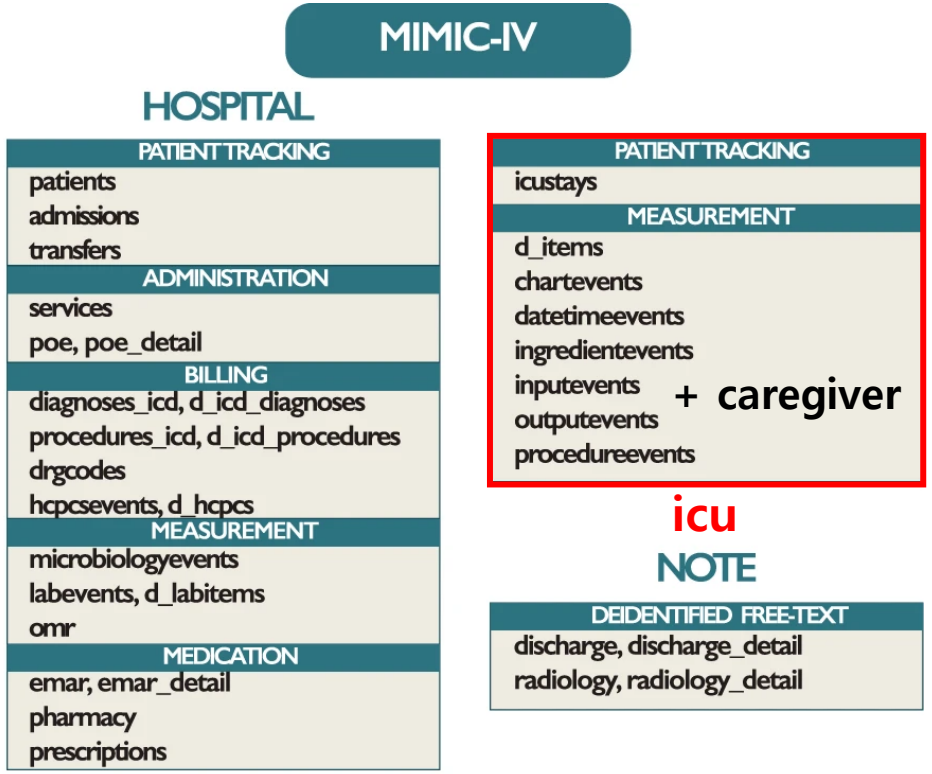
Data description - *hosp* module

A single patient's hospitalization (hadm_id 28503629)



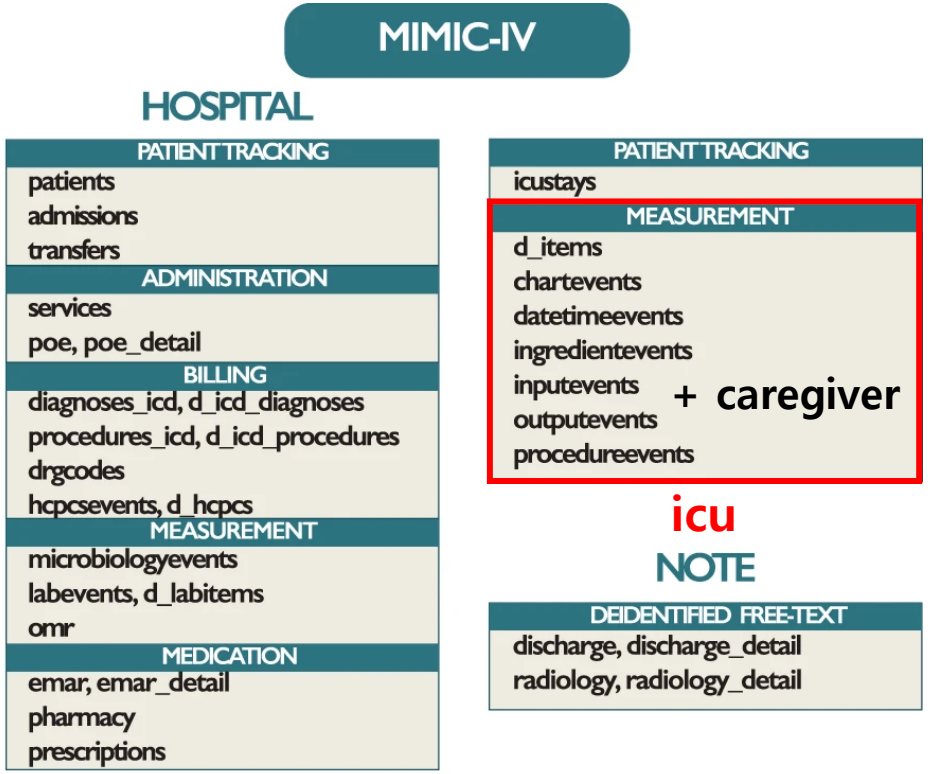


Data description - *icu* module



- 94,458 ICU stays
- 65,366 unique patients
- **Sources**: MetaVision
- **Identifiers**
  - subject_id
  - stay_id : ICU stay
    - Consecutive transfers => a single stay_id
    - Transfer to a non-ICU ward between two ICU stays -> unique stay_id for each stay
  - itemid : identification of the concept in d_items
  - ***icustays*** : records of ICU stays
    - Derived from the ***transfers*** table in the *hosp* module

Data description - *icu* module



- **Measurement**
  - "events" based on the data type
  - ***inputevents*** : intravenous and fluid inputs
  - ***ingredientevents*** : ingredients for the inputs
  - ***outputevents*** : patient outputs
  - ***procedureevents*** : procedures including organ support treatments
  - ***datetimeevents*** : information documented as a date or time
  - ***chartevents*** : other charted information at the bedside

MIMIC-IV Usage note

SNUH

- Available in PhysioNet

Database | Credentialed Access

MIMIC-IV

Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Brian Gow, Benjamin Moody, Steven Horng, Leo Anthony Celi, Roger Mark

Published: July 23, 2024. Version: 3.0

Guidelines for creating datasets and models from MIMIC *(April 24, 2024, 10:12 a.m.)*

We recognize that there is value in creating datasets or models that are either derived from MIMIC or which augment MIMIC in some way (for example, by adding annotations). Here are some guidelines on creating these datasets and models:

- **Any derived datasets or models should be treated as containing sensitive information**. If you wish to share these resources, they should be shared on PhysioNet under the same agreement as the source data.
- **If you would like to use the MIMIC acronym in your project name**, please include the letters "Ext" (for example, MIMIC-IV-Ext-YOUR-DATASET"). Ext may either indicate "extracted" (e.g. a derived subset) or "extended" (e.g. annotations), depending on your use case.

When using this resource, please cite: (show more options)
Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L. A., & Mark, R. (2024). MIMIC-IV (version 3.0). *PhysioNet*. https://doi.org/10.13026/hxp0-hg59.

Additionally, please cite the original publication:
Johnson, A.E.W., Bulgarelli, L., Shen, L. et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 10, 1 (2023). https://doi.org/10.1038/s41597-022-01899-x

Please include the standard citation for PhysioNet: (show more options)
Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.

Contents

Share

Access

Access Policy:
Only credentialed users who sign the DUA can access the files.

License (for files):
PhysioNet Credentialed Health Data License 1.5.0

Data Use Agreement:
PhysioNet Credentialed Health Data Use Agreement 1.5.0

Required training:
CITI Data or Specimens Only Research

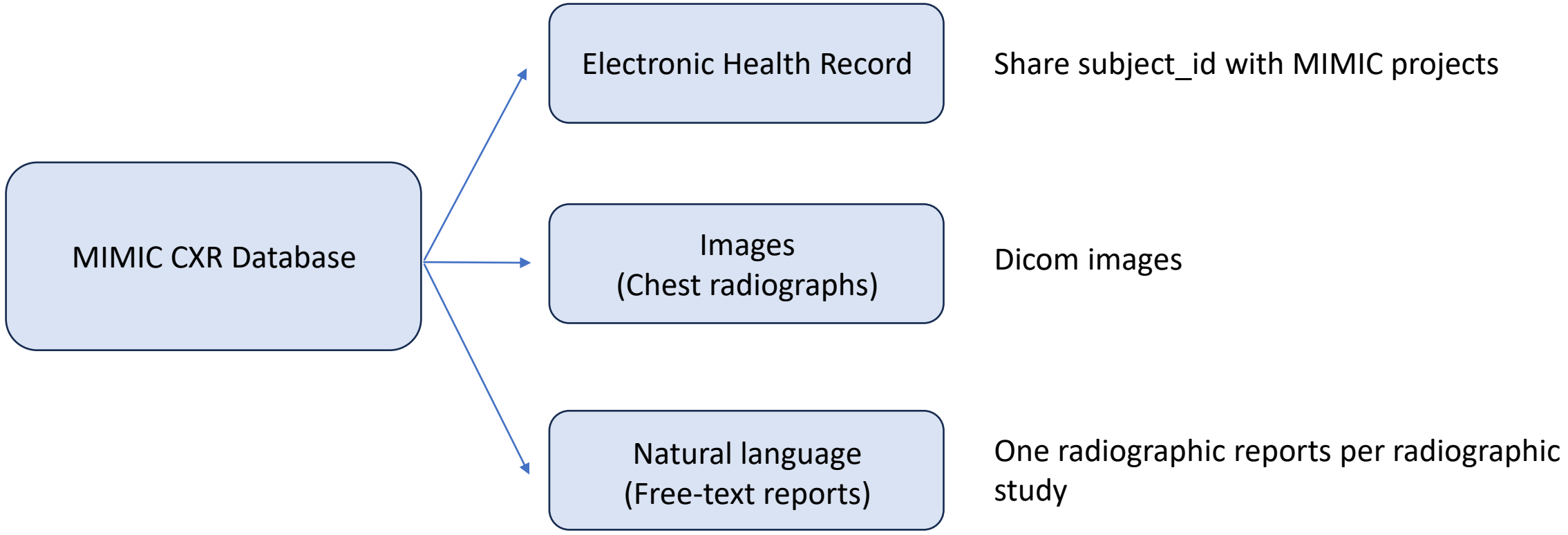
Abstract

Retrospectively collected medical data has the opportunity to improve patient care through knowledge discovery and algorithm development. Broad reuse of medical data is desirable for the greatest public good, but data sharing must be done in a manner which protects patient privacy. Here we present Medical Information Mart for Intensive Care (MIMIC)-IV, a large deidentified dataset of patients admitted to the emergency department or an intensive care unit at the Beth Israel Deaconess Medical Center in Boston, MA. MIMIC-IV contains data for over 65,000 patients admitted to an ICU and over 200,000 patients admitted to the emergency department. MIMIC-IV incorporates contemporary data and adopts a modular approach to data organization, highlighting data provenance and facilitating both individual and combined use of disparate data sources. MIMIC-IV is intended to carry on the success of MIMIC-III and support a broad set of applications within healthcare.

CONTENTS

MIMIC-CXR

- MIMIC-CXR is only available between 2011 - 2016 for patients who were admitted to the emergency department, and this selection bias impacts the patient cohort.
- Large dataset (~4.7TB) with 377,110 images corresponding to 227,835 radiographic studies from 65,079 patients



Files

Folders	10 folders (p10 - p19), each with ~6,500 sub-folders. Sub-folders are named according to th e patient identifier, and contain free-text reports and DICOM files for all studies for that pati ent
CXR-record-list.csv.gz	a compressed file providing the link between an image, its corresponding study identifier, a nd its corresponding patient identifier
cxr-study-list.csv.gz	a compressed file providing a link between anonymous study and patient identifiers
cxr-provider-list.csv.gz	a compressed file providing the ordering, attending, and resident provider associated with t he given radiology study
mimic-cxr-reports.tar.gz	for convenience, all free-text reports have been compressed in a single archive file

Folders & files

Folders	10 folders (p10 - p19), each with ~6,500 sub-folders. Sub-folders are named according to the patient identifier, and contain free-text reports and DICOM files for all studies for that patient
CXR-record-list.csv.gz	a compressed file providing the link between an image, its corresponding study identifier, and its corresponding patient identifier
cxr-study-list.csv.gz	a compressed file providing a link between anonymous study and patient identifiers
cxr-provider-list.csv.gz	a compressed file providing the ordering, attending, and resident provider associated with the given radiology study
mimic-cxr-reports.tar.gz	for convenience, all free-text reports have been compressed in a single archive file

```
files
└── p10
    └── p10000032
        ├── s50414267
        │   ├── 02aa804e-bde0afdd-112c0b34-7bc16630-4e384014.dcm
        │   └── 174413ec-4ec4c1f7-34ea26b7-c5f994f8-79ef1962.dcm
        ├── s50414267.txt
        ├── s53189527
        │   ├── 2a2277a9-b0ded155-c0de8eb9-c124d10e-82c5caab.dcm
        │   └── e084de3b-be89b11e-20fe3f9f-9c8d8dfe-4cfd202c.dcm
        ├── s53189527.txt
        ├── s53911762
        │   ├── 68b5c4b1-227d0485-9cc38c3f-7b84ab51-4b472714.dcm
        │   └── fffabebf-74fd3a1f-673b6b41-96ec0ac9-2ab69818.dcm
        ├── s53911762.txt
        ├── s56699142
        │   └── ea030e7a-2e3b1346-bc518786-7a8fd698-f673b44c.dcm
        └── s56699142.txt
```

Images & reports

- Personal Health Information (PHI) were both removed from both the DICOM meta-data and the pixel values
- Images sometimes contain "burned in" annotations : removed by custom algorithms except orientation information

Fig. 1

EXAMINATION: CHEST (PA AND LAT)

INDICATION: ___ year old woman with ?pleural effusion // ?pleural effusion

TECHNIQUE: Chest PA and lateral

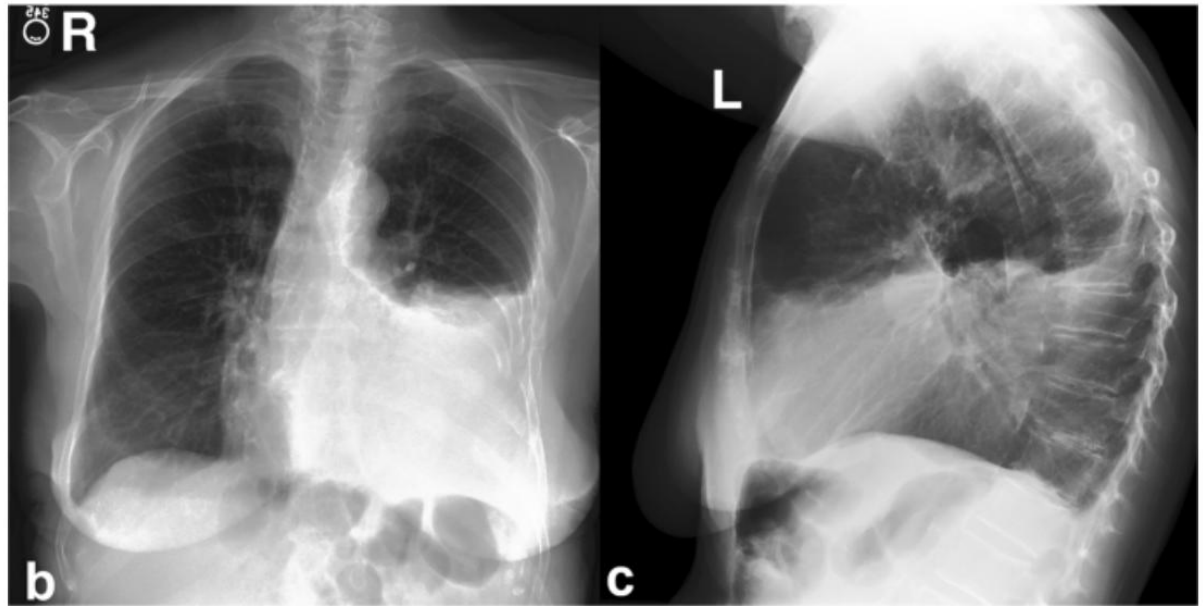
COMPARISON: ___

FINDINGS: FINDINGS

Cardiac size cannot be evaluated. Large left pleural effusion is new. Small right effusion is new. The upper lungs are clear. Right lower lobe opacities are better seen in prior CT. There is no pneumothorax. There are mild degenerative changes in the thoracic spine

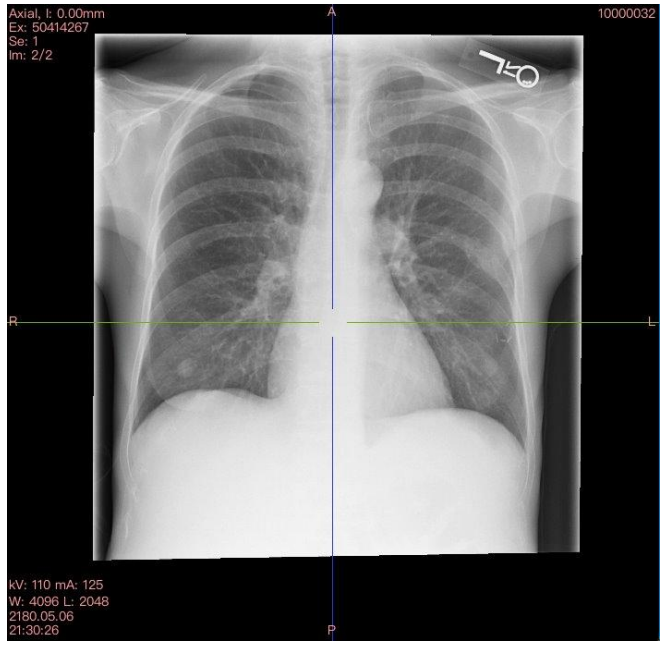
IMPRESSION: IMPRESSION

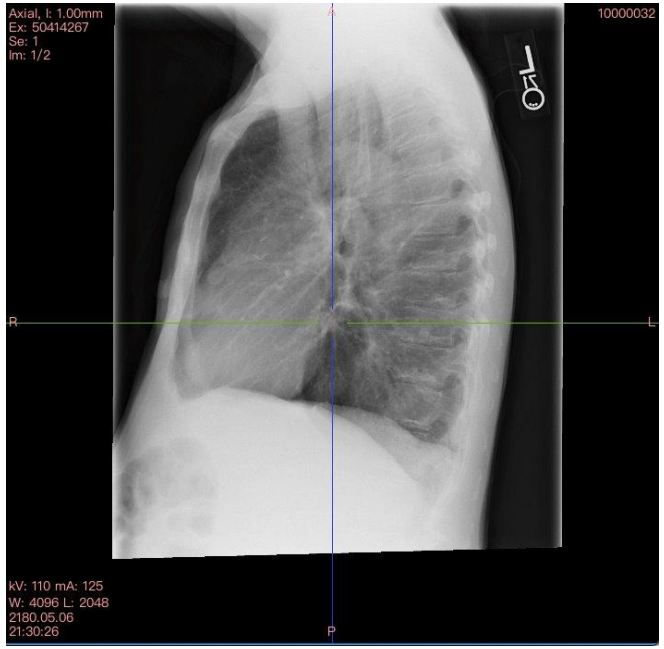
a Large left pleural effusion



Example study contained in MIMIC-CXR. Above (**a**), the radiology report provides the interpretation of the image. PHI has been removed and replaced with three underscores (_ _ _). Below, the two chest radiographs for this study are shown: (**b**) the frontal view (left image) and (**c**) the lateral view (right image).

Images & reports





Subject ID : p10000032
Study ID : s50414267

FINAL REPORT EXAMINATION: CHEST (PA AND LAT) INDICATION: ___F with new onset ascites // eval for infection TECHNIQUE: Chest PA and lateral COMPARISON: None. FINDINGS: There is no focal consolidation, pleural effusion or pneumothorax. Bilateral nodular opacities that most likely represent nipple shadows. The cardiomediastinal silhouette is normal. Clips project over the left lung, potentially within the breast. The imaged upper abdomen is unremarkable. Chronic deformity of the posterior left sixth and seventh ribs are noted. IMPRESSION: No acute cardiopulmonary process.

MIMIC-CXR-JPG

- **DICOM files** : contain higher pixel depth that can be perceived by the human eye (16bit)
- **JPG** : compressed image file format with 12bit
- **Total 570.3GB** (Original MIMIC-CXR : 4.3TB)

DICOM CXR

Preprocessing

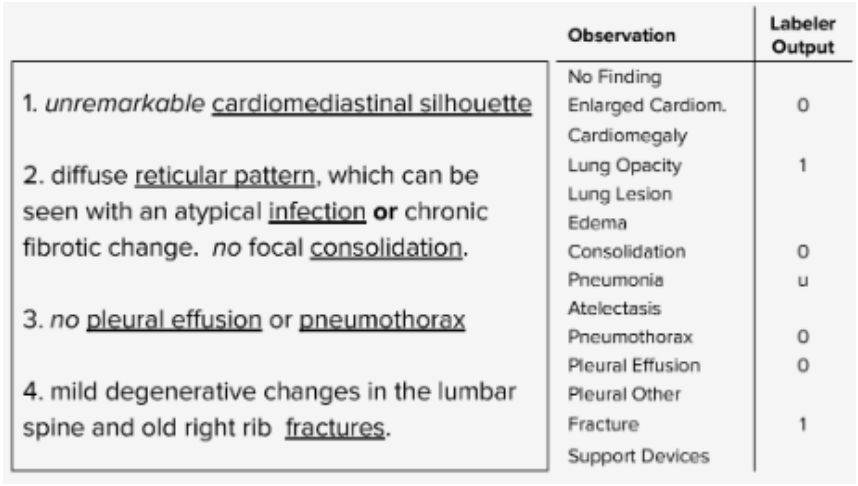
Compressed JPG format

Radiologic Reports (Free Text)

CheXpert Labeler

Structured Labels

1. Atelectasis
2. Cardiomegly
3. Consolidation
4. Edema
5. Enlarged cardiomediastinum
6. Fracture
7. Lung Lesion
8. Lung Opacity
9. Pleural Effusion
10. Pneumothorax
11. Pleural other
12. Support devices
13. No Finding



1. *unremarkable* cardiomediastinal silhouette

2. diffuse reticular pattern, which can be seen with an atypical infection **or** chronic fibrotic change. *no* focal consolidation.

3. *no* pleural effusion or pneumothorax

4. mild degenerative changes in the lumbar spine and old right rib fractures.

Observation	Labeler Output
No Finding	
Enlarged Cardiom.	0
Cardiomegaly	
Lung Opacity	1
Lung Lesion	
Edema	
Consolidation	0
Pneumonia	u
Atelectasis	
Pneumothorax	0
Pleural Effusion	0
Pleural Other	
Fracture	1
Support Devices	

CONTENTS

MIMIC-IV-NOTE

- MIMIC-IV-NOTE is a large, open access database of deidentified free-text clinical notes for patients included in the MIMIC-IV clinical database (ICU patients) (Released ‘2023.01).
- Notes written within one year of a patient encounter(ER visit or hospital stay) are included.

Deidentification

The combination of “custom rule-based approach” and “neural network” was used to deidentify the notes.

- PII was removed in accordance with Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor provision.
- PII was replaced with three underscores (___).

Custom rule-based approach

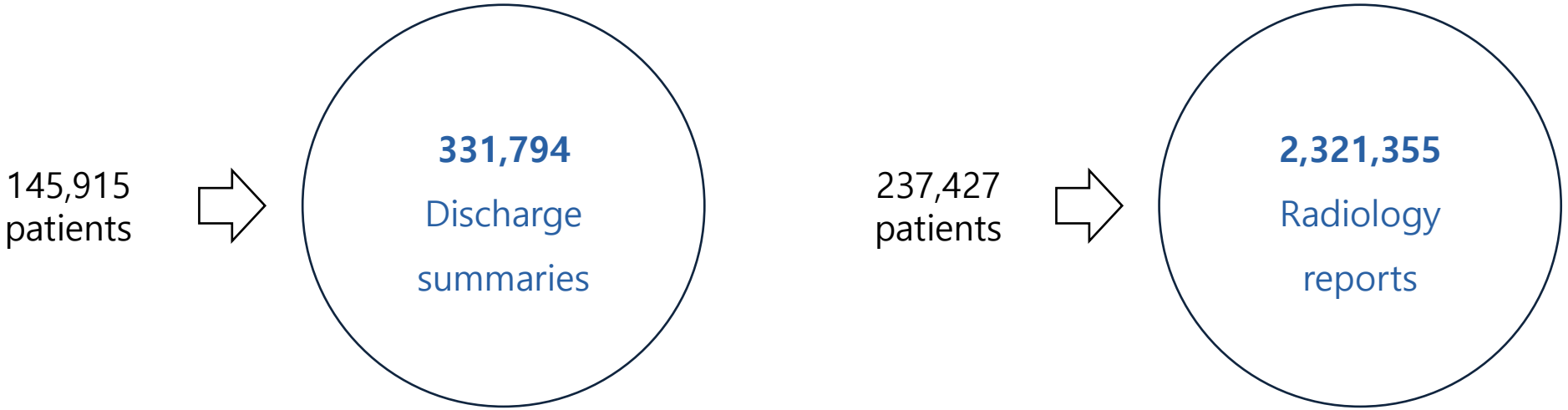
- uses predefined rules and patterns to identify and remove PII.
  - text patterns, keywords, identifiable sequences
- precise and reliable for well-defined patterns

Neural network

- Neural network, trained by a dataset of annotated text with sensitive information already tagged, automatically identify and remove PII.
- flexible and can generalize better

Note contents

- MIMIC-IV-NOTE comprises four tables: discharge, discharge_detail, radiology, and radiology_Detail
- **Discharge** : 331,794 deidentified discharge summaries from 145,915 patients (admitted to the hospital and emergency department at the Beth Israel Deaconess Medical Center in Boston, MA, USA)
- **Radiology** : 2,321,355 deidentified radiology reports for 237,427 patients



Discharge data

Content #1: Discharge data includes discharge summaries and associated information.

- **“Discharge” table** contains discharge summary, which is a data on the reason for admission, hospital course, and relevant discharge instructions.

	A	B	C	D	E	F	G	H	I
1		note_id	subject_id	hadm_id	note_type	note_seq	charttime	storetime	text
	0	10000032-DS-21	10000032	22595853	DS	21	2180-05-07 0:00	2180-05-09 15:26	₩nName: __ Unit No: __₩n ₩nAdmission Date: __ Discharge Date: __₩n ₩nDate of Birth: __ Sex: F₩n ₩nService: MEDICINE₩n ₩nAllergies: ₩nNo Known Allergies / Adverse D Procedure:₩nParacentesis₩n₩n ₩nHistory of Present Illness:₩n__ HCV cirrhosis c/b ascites, hiv on ART, h/o IVDU, COPD, ₩nbioplar, PTSD, presented from OSH ED with worsenir anything" and that ₩nshe "doesn't want to put more chemicals in her." She does not ₩nfollow Na-restricted diets. In the past week, she notes that she ₩nhas been having worsen from eating stale ₩ncake (n/v 20 min after food ingestion), which resolved the same ₩nday. She denies other recent illness or sick contacts. She notes ₩nthat she has been noticir of her abd pain, she went to OSH ED and was transferred ₩nto __ for further care. Per ED report, pt has brief period of ₩nconfusion - she did not recall the ultrasound or bloodw were 98.4 70 106/63 16 97%RA ₩nLabs notable for ALT/AST/AP __ __: __, ₩nTbili1.6, WBC 5K, platelet 77, INR 1.6 ₩n₩n ₩nPast Medical History:₩n1. HCV Cirrhosis ₩n2. No l disease, she is being followed by Dr. __ Dr. ₩n__. ₩n5. COPD ₩n6. Past history of smoking. ₩n7. She also had a skin lesion, which was biopsied and showed ₩nskin cancer per p ₩ndiscoloration. It was biopsied to exclude the possibility of ₩n__'s sarcoma, the results is pending. ₩n9. A 15 mm hypoechoic lesion on her ultrasound on __ ₩nand is being m cocaine and heroin use. ₩n₩n ₩nSocial History:₩n__₩nFamily History:₩nShe a total of five siblings, but she is not ₩ntalking to most of them. She only has one brother that she two months ago. No ₩nregular alcohol consumption. Last drug use __ years ago. She ₩nquit smoking a couple of years ago. ₩n₩n ₩nPhysical Exam:₩nVS: 98.1 107/61 78 18 9 expiratory phase, no w/r/r ₩nAbdomen: distended, mild diffuse tenderness, +flank dullness, ₩ncannot percuss liver/spleen edge __ distension ₩nGU: no foley ₩nExt: wwp, no c/e, ₩nVS: 98 105/70 95₩nGeneral: in NAD ₩nHEENT: anicteric sclera, OP clear ₩nNeck: supple, no LAD ₩nCV: RRR,S1S2, no m/r/g ₩nLungs: CTAb, prolonged expiratory phase, no ₩nPertinent Results:₩n__ 10:25PM GLUCOSE-109* UREA N-25* CREAT-0.3* SODIUM-138 ₩nPOTASSIUM-3.4 CHLORIDE-105 TOTAL CO2-27 ANION GAP-9₩n__ 10:25PM estGl 3.3*₩n__ 10:25PM WBC-5.0# RBC-4.29 HGB-14.3 HCT-42.6 MCV-99* ₩nMCH-33.3* MCHC-33.5 RDW-15.7*₩n__ 10:25PM NEUTS-70.3* LYMPHS-16.5* MONOS-8.1 EOS-4.2 ₩nU/S: ₩n1. Nodular appearance of the liver compatible with cirrhosis. ₩nSigns of portal ₩nhypertension including small amount of ascites and splenomegaly. ₩n ₩n2. Cholelith and discomfort.₩n ₩nBrief Hospital Course:₩n__ HCV cirrhosis c/b ascites, hiv on ART, h/o IVDU, COPD, ₩nbioplar, PTSD, presented from OSH ED with worsening abd ₩ndisten

Discharge data

Content #1: Discharge data includes discharge summaries and associated information.

- **"Discharge_detail" table** contains auxillary information related to the discharge summary. (deidentified placeholder for the authors)

	A	B	C	D	E
1	note_id	subject_id	field_name	field_value	field_ordinal
2	10000032-DS-21	10000032	author	___	1
3	10000032-DS-22	10000032	author	___	1
4	10000032-DS-23	10000032	author	___	1
5	10000032-DS-24	10000032	author	___	1
6	10000084-DS-17	10000084	author	___	1
7	10000117-DS-21	10000117	author	___	1
8	10000117-DS-22	10000117	author	___	1
9	10000248-DS-10	10000248	author	___	1
10	10000560-DS-15	10000560	author	___	1
11	10000764-DS-11	10000764	author	___	1
12	10000826-DS-17	10000826	author	___	1
13	10000826-DS-18	10000826	author	___	1
14	10000826-DS-19	10000826	author	___	1

Radiology data

Content #2: Radiology data includes unstructured radiology reports and the information associated with the imaging study.

- **"Radiology" table** contains semi-structured radiology reports, which follows a certain template for a given imaging protocol.

	A	B	C	D	E	F	G	H	I
1		note_id	subject_id	hadm_id	note_type	note_seq	charttime	storetime	text
2	0	10000032-RR-14	10000032	22595853	RR	14	##########	##########	EXAMINATION: CHEST (PA AND LAT)₩n₩nINDICATION: ___ with new onset ascites // eval for infection₩n₩nTECHNIQUE: Chest PA and pneumothorax. Bilateral₩nnodular opacities that most likely represent nipple shadows. The₩ncardiomediastinal silhouette is normal. Cli unremarkable.₩nChronic deformity of the posterior left sixth and seventh ribs are noted.₩n₩nIMPRESSION: ₩n₩nNo acute cardiopulm
3	1	10000032-RR-15	10000032	22595853	RR	15	##########	##########	EXAMINATION: LIVER OR GALLBLADDER US (SINGLE ORGAN)₩n₩nINDICATION: ___ year-old female with cirrhosis, jaundice.₩n₩nTECH were₩nobtained.₩n₩nCOMPARISON: None.₩n₩nFINDINGS: ₩n₩nLIVER: The liver is coarsened and nodular in echotexture. There is no main hepatic artery shows normal arterial waveform. ₩nThere is a small amount of ascites.₩n₩nBILE DUCTS: There is no intrahepatic bi gallstone.₩n₩nPANCREAS: Imaged portion of the pancreas appears within normal limits, without₩nmasses or pancreatic ductal dilation measuring 13.5 cm.₩n₩nKIDNEYS: The right kidney measures 12.1 cm. The left kidney measures 13.4 cm.₩nNormal cortical echogenicity aorta and IVC are within normal₩nlimits.₩n₩nIMPRESSION: ₩n₩n1. Nodular appearance of the liver compatible with cirrhosis. Signs of portal veins with normal hepatopetal flow.₩n
4	2	10000032-RR-16	10000032	22595853	RR	16	2180-05-07 9:55	##########	INDICATION: ___ HCV cirrhosis c/b ascites, hiv on ART, h/o IVDU, COPD,₩nbioplar, PTSD, presented from OSH ED with worsening abd c paracentesis₩n₩nCOMPARISON: Abdominal ultrasound ___₩n₩nFINDINGS: ₩n₩nLimited grayscale ultrasound imaging of the abdome was selected for paracentesis.₩n₩nPROCEDURE: The procedure, risks, benefits and alternatives were discussed₩nwith the patient and w procedure,₩nconfirming the patient's identity with 3 identifiers, and reviewing a₩nchecklist per ___ protocol.₩n₩nUnder ultrasound gui lidocaine was instilled₩nfor local anesthesia.₩n₩nA 5 ___ catheter was advanced into the largest fluid pocket in the right₩nlower quad immediate complication.₩nEstimated blood loss was minimal. A sample of the fluid was sent to the lab₩nas requested.₩n₩nDr. ___ atte ₩n₩nSuccessful uncomplicated ultrasound guided diagnostic and therapeutic₩nparacentesis yielding 1.5 L of serosanguineous fluid fr

Radiology data

Content #2: Radiology data includes unstructured radiology reports and the information associated with the imaging study.

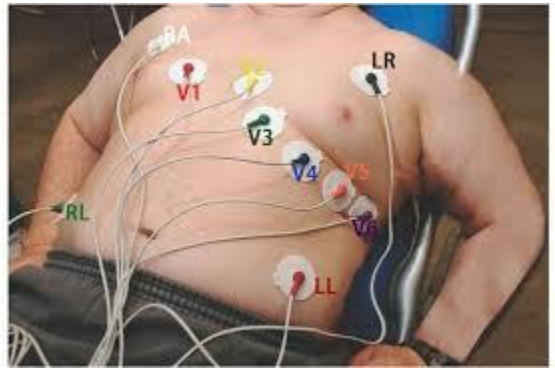
- **"Radiology_detail" table** contains auxillary information related to the imaging study.

	A	B	C	D	E
1	note_id	subject_id	field_name	field_value	field_ordinal
2	10000032-RR-14	10000032	exam_code	C11	1
3	10000032-RR-14	10000032	exam_name	CHEST (PA & LAT)	1
4	10000032-RR-15	10000032	exam_code	U314	1
5	10000032-RR-15	10000032	exam_code	U644	3
6	10000032-RR-15	10000032	exam_code	W82	2
7	10000032-RR-15	10000032	exam_name	-59 DISTINCT PROCEDURAL SERVICE	2
8	10000032-RR-15	10000032	exam_name	DUPLEX DOP ABD/PEL LIMITED	3
9	10000032-RR-15	10000032	exam_name	LIVER OR GALLBLADDER US (SINGLE ORGAN)	1
10	10000032-RR-16	10000032	exam_code	U321	1
11	10000032-RR-16	10000032	exam_name	PARACENTESIS DIAG/THERAP W IMAGING GUID	1

CONTENTS

MIMIC-IV-ECG

- Background
  - Routinely obtained
  - 12 lead ECG vs. Continuous monitoring
  - I, II, III, aVF, aVR, aVL, V1, V2, V3, V4, V5, V6
  - MIMIC-IV Clinical Database patients



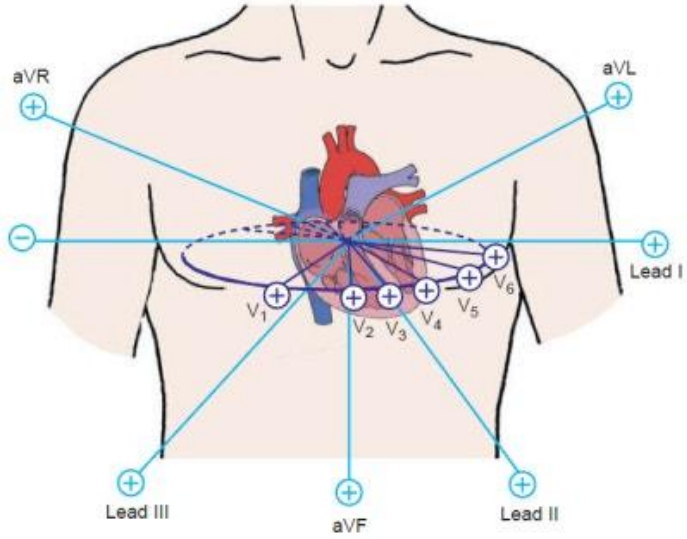
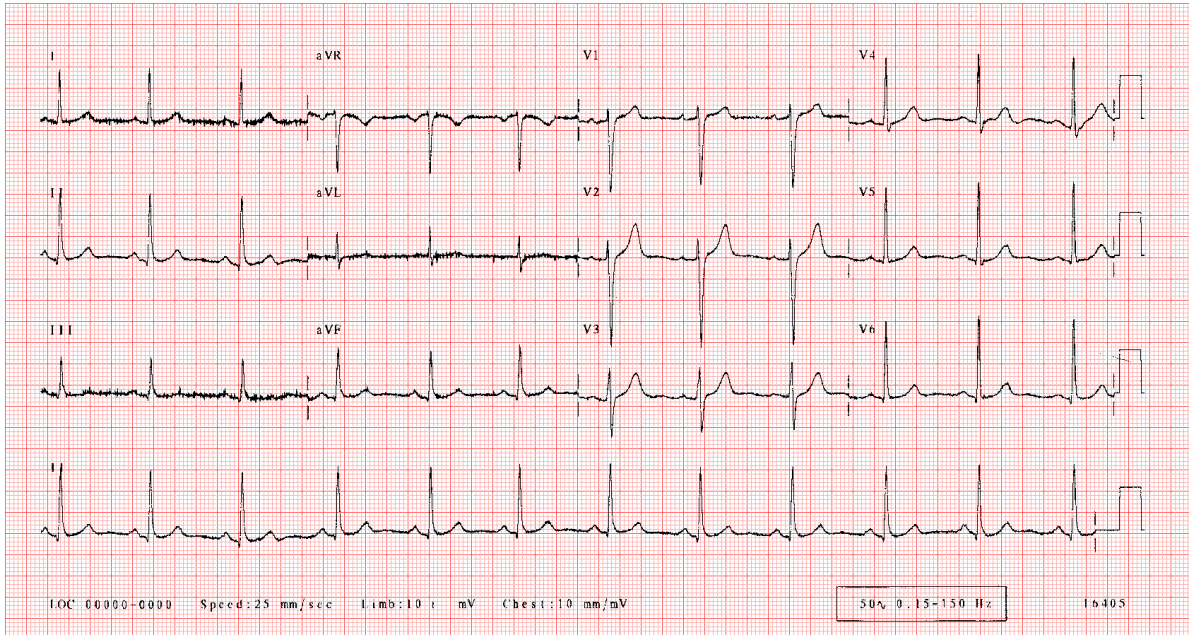


FIGURE 17-32 ▲ Electrocardiographic views of the heart.



Waveforms

- subject-id matched with MIMIC-IV / date-times shifted / study_id

1. Electrocardiogram Waveforms
   - Emerency department, ward, ICU, OPD – WFDB format
   - Header (.hea) and Signal (.dat) / signal, subject_id, shifted date-time
   - Timestamps aligned with MIMIC-IV clinical database *

   - ~160,000 subjects / ~ 800,000 ECGs / 10s long / sampled at 500 Hz
   - 5% saved for future use
   - ~55% admission / ~25% ED

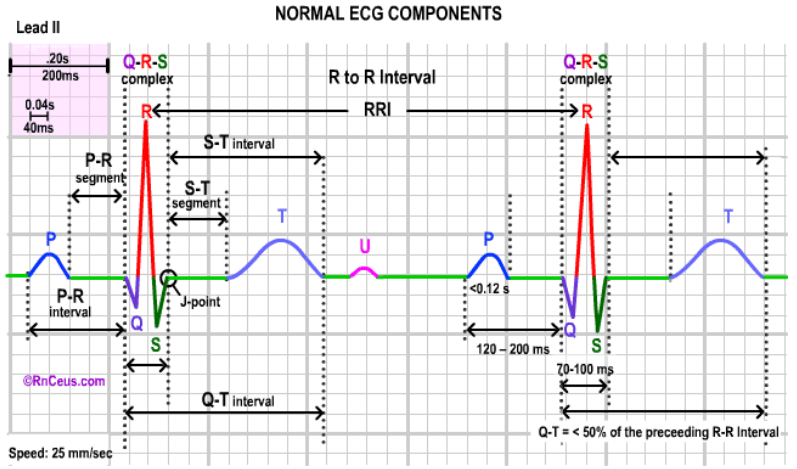
   - record_list.csv

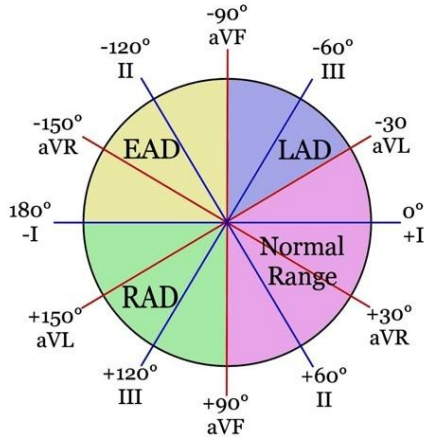
   * Time stamps may be out of sync

```
files
├── p1000
│   └── p10001725
│       └── s41420867
│           ├── 41420867.dat
│           └── 41420867.hea
└── p1002
    └── p10023771
        ├── s42745010
        │   ├── 42745010.dat
        │   └── 42745010.hea
        ├── s46989724
        │   ├── 46989724.dat
        │   └── 46989724.hea
        └── s42460255
            ├── 42460255.dat
            └── 42460255.hea
```

Machine measurements

2. Machine measurements





	A	B	C	D	E	F	G	H	I	J	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	subject_id	study_id	cart_id	ecg_time	report_0	report_1	report_2	report_3	report_4	report_5	bandwidth	filtering	rr_interval	p_onset	p_end	qrs_onset	qrs_end	t_end	p_axis	qrs_axis	t_axis
2	10000032	40689238	6848296	2180-07-23 8:44	Sinus rhyth	Possible right atrial ab	Borderline ECG				0.005-150 Hz	60 Hz notch Baseline filter	659	40	128	170	258	518	81	77	79
3	10000032	44458630	6848296	2180-07-23 9:54	Sinus rhyth	Possible right atrial ab	Borderline ECG				0.005-150 Hz	60 Hz notch Baseline filter	722	40	124	162	246	504	77	75	70
4	10000032	49036311	6376932	2180-08-06 9:07	Sinus tachycardia	Normal ECG except for rate					0.005-150 Hz	60 Hz notch Baseline filter	600	40	130	162	244	474	79	72	77
5	10000117	45090959	6214760	2181-03-04 17:14	Sinus rhythm	Normal ECG					0.005-150 Hz	60 Hz notch Baseline filter	659	40	146	180	254	538	79	66	69
6	10000117	48446569	6632385	2183-09-18 13:52	Sinus rhythm						0.0005-150 Hz	<not specified>	659	368	29999	504	590	868	84	80	77
7	10000285	42709053	6632385	2159-11-26 14:29	Sinus rhythm						0.0005-150 Hz	<not specified>	822	365	29999	499	592	852	26	46	30
8	10000560	41445586	6852956	2189-10-03 12:54	Sinus rhythm	Normal ECG					0.005-150 Hz	60 Hz notch Baseline filter	952	40	146	198	282	598	24	80	20
9	10000560	42695383	6551957	2198-09-19 10:01	Sinus rhyth	rSr'(V1) - p	Low QRS voltages in p	Borderline ECG			0.005-150 Hz	60 Hz notch Baseline filter	923	40	140	188	278	594	26	86	13
10	10000635	40067704	6924910	2136-06-19 7:24	Sinus rhythm with PAC	Borderline ECG					0.005-150 Hz	60 Hz notch Baseline filter	952	40	180	196	294	610	59	-17	3
11	10000635	45386375	6919786	2136-07-05 12:35	Sinus rhyth	Extensive T wave chan	Abnormal ECG				0.005-150 Hz	60 Hz notch Baseline filter	1000	40	156	178	274	584	8	-11	19
12	10000635	43522917	6919786	2137-02-07 13:58	Sinus brady	Left ventric	Extensive T wave chan	Abnormal ECG			0.005-150 Hz	60 Hz notch Baseline filter	1052	40	138	172	268	624	55	-16	2

Cardiologist reports

3. Cardiologist Reports
   - ~ 600,000 reports
   - free-text note in the MIMIC-IV-Note

   - waveform_note_links.csv
   - note_id

	A	B	C	D	E	F
1	subject_id	study_id	waveform_path	note_id	note_seq	charttime
2	10000032	40689238	files/p1000/p10000032/s40689238/40689238	10000032-EK-4	4	2180-07-23 8:44
3	10000032	44458630	files/p1000/p10000032/s44458630/44458630	10000032-EK-5	5	2180-07-23 9:54
4	10000032	49036311	files/p1000/p10000032/s49036311/49036311	10000032-EK-6	6	2180-08-06 9:07
5	10000117	48446569	files/p1000/p10000117/s48446569/48446569	10000117-EK-17	17	2183-09-18 13:52
6	10000560	41445586	files/p1000/p10000560/s41445586/41445586	10000560-EK-12	12	2189-10-03 12:54
7	10000635	40067704	files/p1000/p10000635/s40067704/40067704	10000635-EK-12	12	2136-06-19 7:24
8	10000635	48339811	files/p1000/p10000635/s48339811/48339811	10000635-EK-13	13	2136-06-20 8:54
9	10000635	45386375	files/p1000/p10000635/s45386375/45386375	10000635-EK-14	14	2136-07-05 12:35
10	10000635	43522917	files/p1000/p10000635/s43522917/43522917	10000635-EK-15	15	2137-02-07 13:58
11	10000635	44095784	files/p1000/p10000635/s44095784/44095784	10000635-EK-16	16	2138-09-29 11:09
12	10000635	42947358	files/p1000/p10000635/s42947358/42947358	10000635-EK-17	17	2139-01-22 14:09
13	10000764	47218930	files/p1000/p10000764/s47218930/47218930	10000764-EK-19	19	2132-10-14 19:40

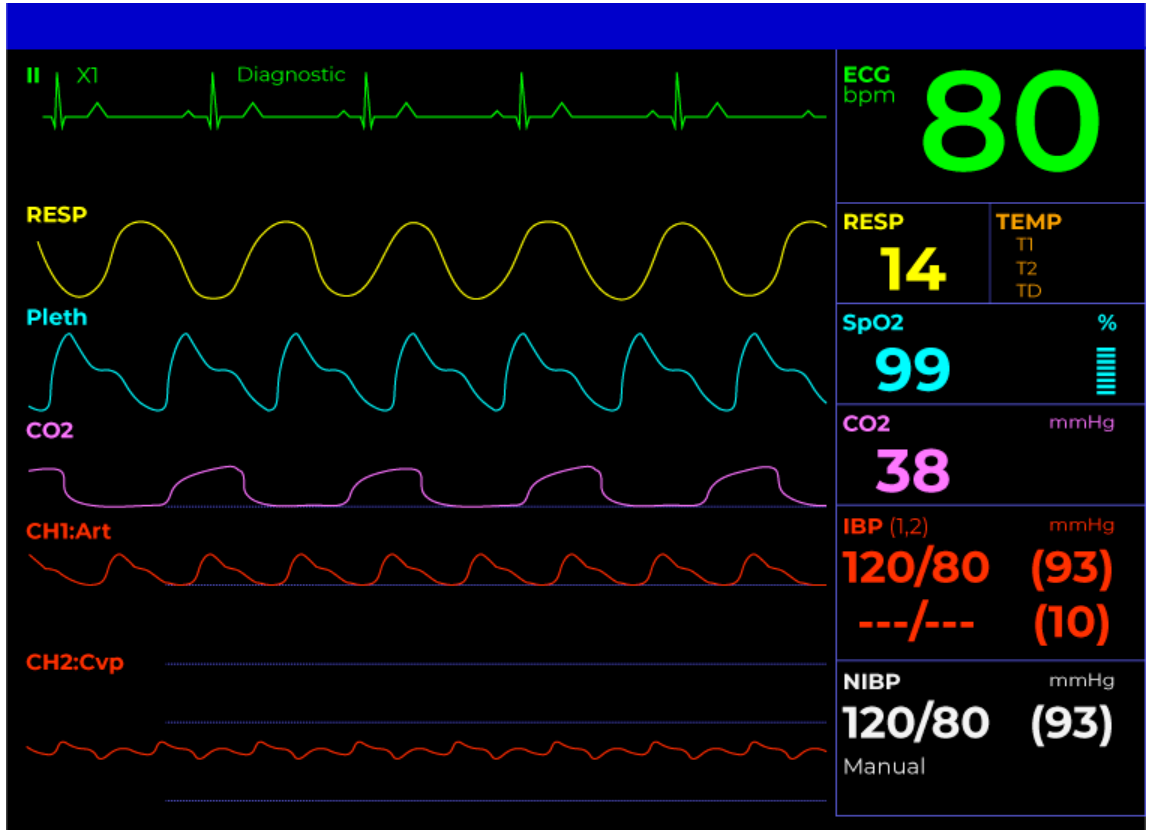
- BigQuery
  - record_list.csv / machine_measurements.csv / waveform_note_links.csv

MIMIC-IV Waveform

- Background
  - Real-time data from ICU bedside monitor
  - MIMIC-IV clinical database – ICU patients
  - 200 records from 198 patients

Data format

- Methods & Data Description
  - Waveform data – WFDB format
    - High-resolution, regularly sampled time series
  - Numeric data – CSV format
    - Digitally derived (e.g. ST, QTc)
    - Sampled irregularly (e.g. NIBP)
  - De-identification
    - Subject_id / hadm_id / shifted date



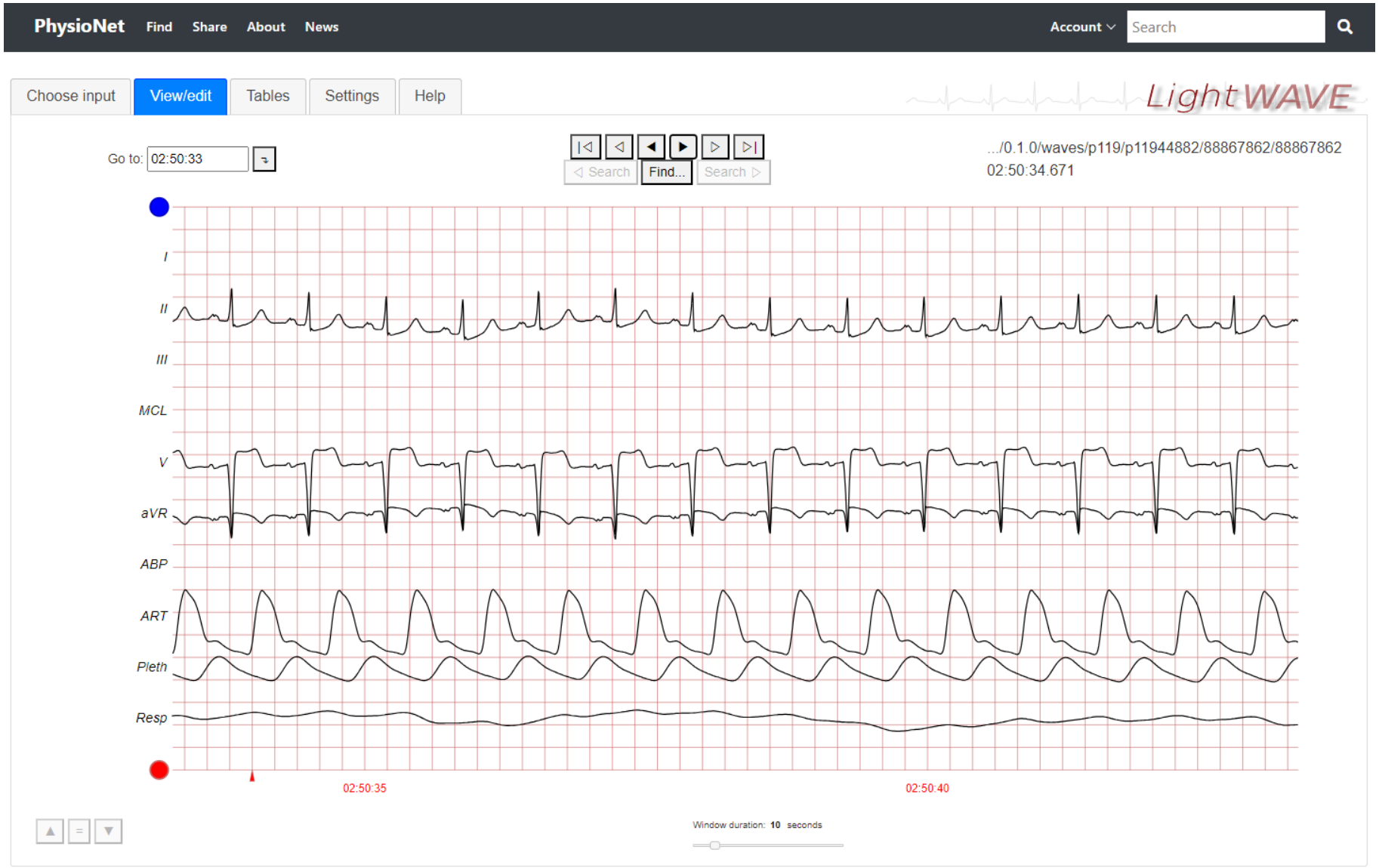
Data structure

- p(subject_id) > (hadm_id) >
  - multi-segment header file (hadm_id).hea
    - general information about the record (starting date, time, length, subject and hospital admission IDs) and the list of waveform segments
  - layout header file (hadm_id)_0000.hea
    - list of waveform signals present in the record
  - One or more segment header files (hadm_id)_0001.hea
  - One or more signal files (hadm_id)_0001e.dat
  - Numerics file (hadm_id)n.csv.gz

Recording

- Record splitting
  - if >1h absence of data
  - may have multiple study in 1 ICU stay
- Storage formats
  - waveform data > flac
  - numeric data > dictzip (.gz)

Visualization tool



Numeric data

- CVP, HR(ECG, SpO2), NBP, PAP, QTc, RR(ECG, aw), ST analysis, SpO2, etCO2

A399742 | 121850880

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	time	CVPm [mr	DeltaQTc [	HR [bpm]	IPI [NU]	NBPd [mm	NBPm [mr	NBPs [mm	PAPd [mm	PAPm [mn	PAPs [mm	PVC [/min	Pulse (NBP	Pulse (PAP	Pulse (SpO	QT [msec]	QT-HR [bp	QTc [msec	RR [rpm]	ST-III [mm	ST-V [mm]	SpO2 [%]	awRR [rpm	btbHR [bpi	etCO2 [mr	imCO2 [mmHg]				
399712	1.22E+08																							147.6						
399713	1.22E+08	-2	71	148								8			148	312	147	488	20	-0.4	0.3	96.4								
399714	1.22E+08																							147.42						
399715	1.22E+08																							147.42						
399716	1.22E+08																							147.6						
399717	1.22E+08	-2	71	148								8			147	312	147	488	18	-0.4	0.3	96.5								
399718	1.22E+08																							147.97						
399719	1.22E+08																							147.6						
399720	1.22E+08	-1	71	148								8			147	312	147	488	18	-0.4	0.3	96.7								
399721	1.22E+08																							147.78						
399722	1.22E+08																							147.6						
399723	1.22E+08																							160						
399724	1.22E+08	1	71	148								8			147	312	147	488	18	-0.4	0.3	96.8								
399725	1.22E+08																							147.24						
399726	1.22E+08																							135.9						
399727	1.22E+08	2	71	148								8			147	312	147	488	17	-0.4	0.3	97								
399728	1.22E+08																							159.79						
399729	1.22E+08																							136.05						
399730	1.22E+08	2	71	148								8			146	312	147	488	17	-0.4	0.3	97								
399731	1.22E+08																							158.94						
399732	1.22E+08																							136.99						
399733	1.22E+08																							159.36						
399734	1.22E+08	2	71	148								8			146	312	147	488	17	-0.4	0.3	97								
399735	1.22E+08																							147.97						
399736	1.22E+08	4	71	149								8			145	312	147	488	17	-0.4	0.3	97.1								
399737	1.22E+08																							148.7						
399738	1.22E+08																							144.93						
399739	1.22E+08	5	71	147								8			146	312	147	488	17	-0.4	0.3	97.1								
399740	1.22E+08																							146.88						
399741	1.22E+08																							145.28						
399742	1.22E+08					87	93	101																						
399743	1.22E+08																							146.88						
399744	1.22E+08	10	71	147								8			145	312	147	488	17	-0.4	0.3	97.2								
399745	1.22E+08																							145.45						
399746	1.22E+08																							147.06						
399747	1.22E+08	10	71	148								8			144	312	147	488	17	-0.4	0.3	97.2								
399748	1.22E+08																							147.06						
399749	1.22E+08																							146.7						
399750	1.22E+08																							146.34						
399751	1.22E+08	10	71	148								8			144	312	147	488	16	-0.4	0.3	97.3								
399752	1.22E+08																							146.52						
399753	1.22E+08																							160.43						
399754	1.22E+08	11	71	147								8			142	312	147	488	16	-0.4	0.3	97.3								
399755	1.22E+08																							134.38						

88501826n(waveform예시)

준비 | 접근성: 사용할 수 없음 | 평균: 30462790.25 | 개수: 4 | 합계: 121851161 | 85%

Q&A