APAC Community Call

August 21, 2025

Agenda

- 2025 APAC Studies: Call for Data Partners
- Regional Chapter Mid-year Updates: Australia, Malaysia
- Deep Dive into OHDSI India’s Developments
- 2025 APAC Symposium Updates

Call for Data Partners

2025 APAC Studies

Gastrointestinal Risk of GLP-1 Receptor Agonists versus SGLT-2 and DPP-4 Inhibitors in Type 2 Diabetes: A Multi-Database Observational Study

Yongqi Zheng, Department of Epidemiology and Biostatistics, Peking University
Feng Sun, Department of Epidemiology and Biostatistics, Peking University

Study Background

Pharmacologic Approaches to T2DM Treatment

Metformin: foundational glucose-lowering therapy, typically used as first-line agent
GLP-1RAs: recommended as monotherapy or as add-on therapy to agents such as metformin in patients with T2DM and established ASCVD, CKD, or obesity, based on their cardiovascular and renal benefits

Evaluate Gastrointestinal(GI) Risks of GLP-1RAs Using Real-World Evidence: Rationale & Gap

Safety Concern: Delayed gastric emptying → possible gastroparesis, intestinal obstruction

Conflicting Evidence:

- Several large-scale observational studies (e.g., Sodhi et al., Nielsen et al.) have reported significantly increased risks of gastroparesis and intestinal obstruction
- Several studies (e.g., Gao et al., Ueda et al.) found no significant increase in gastrointestinal obstruction
- There have been no RCTs specifically designed to evaluate the gastrointestinal safety outcomes of GLP-1RAs

Need for Real-World Evidence:

- Prior studies limited by insufficient confounder adjustment
- Few multi-database studies on GI safety of GLP-1RAs

SODHI M, et al. Jama, 2023. NIELSEN J, et al. Inflamm Bowel Dis, 2024. GAO Z, et al. Acta Diabetol, 2025. UEDA P, et al. Clin Gastroenterol Hepatol, 2024.

Study Objectives

- Primary objective
  - Compare the risk of gastroparesis in T2DM patients initiating GLP-1RAs versus DPP-4 inhibitors or SGLT-2 inhibitors
  - Compare the risk of intestinal obstruction in T2DM patients initiating GLP-1RAs versus DPP-4 inhibitors or SGLT-2 inhibitors

- Secondary objective
  - Compare the risk of acute pancreatitis and nonalcoholic fatty liver disease (NAFLD) in T2DM patients initiating GLP-1RAs versus DPP-4 inhibitors or SGLT-2 inhibitors, as secondary outcomes to provide a broader assessment of digestive system safety

Study Design

Analytic use case	Type	Structure
Population-level effect estimation	**Comparative effectiveness**	Does exposure to **GLP-1RAs** have a different risk of experiencing **gastroparesis or intestinal obstruction** within **end of continuous observation**, relative to **DPP4-i or SGLT2is** among the population with **type 2 diabetes and history of metformin**?

Population

Inclusion

- Adults (≥18 years) with T2DM
- ≥365 days prior observation
- ≥90 days prior metformin use

Exclusion

- T1DM or secondary diabetes
- Prior exposure to study drugs or anti-diabetic exposure
- No prior insulin use or combo initiation
- History of pancreatitis, digestive system cancer, or abdominal surgery
- Renal dialysis, renal transplantation or end stage renal disease

Target: GLP-1RAs

Comparator: SGLT-2 inhibitors, DPP-4 inhibitors

Outcome:

Primary outcome:

- Gastroparesis
- Intestinal obstruction

Secondary outcome:

- Acute Pancreatitis
- NAFLD

CohortDiagnostics

- URL of CohortDiagnostics package:
  - https://github.com/ohdsi-studies/2025APACStudy-Peking/tree/master/CohortDiagnostics
- For troubleshooting:
  - Open an issue at https://github.com/ohdsi-studies/2025APACStudy-Peking/issues
- Send your results to:
  - Yongqi Zheng zyq4664@pku.edu.cn
  - OHDSI APAC apacsymposium@ohdsi.org
- Join us at our study Teams channel: 2025 APAC Study 2 - Peking

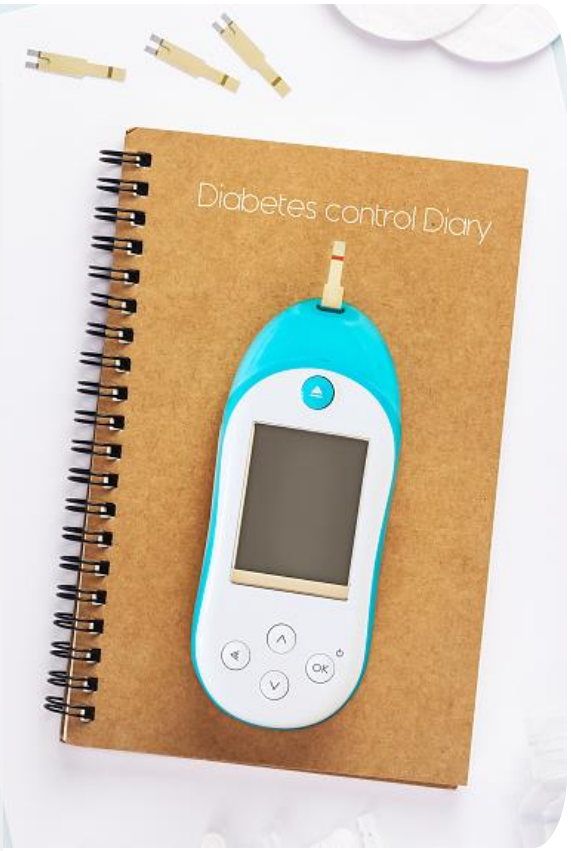
Association Between Fasting Plasma Glucose Levels and Annual Hospitalization Days: A Multicenter Study Using the OHDSI Framework

Ph.D. Jianying Gu, Zhongshan Hospital, Fudan University
Ph.D. Lei Liu, Fudan University

Study Background

Fasting Plasma Glucose Levels

- Abnormal FPG levels relate to metabolic health, diabetes, and increased healthcare resource use.
- High levels can cause disorders, inflammation, and exacerbate chronic diseases, potentially leading to longer hospital stays



OHDSI Framework about FPG: Investigates association between FPG levels and annual hospitalization days using OHDSI framework, to inform glucose management and resource allocation strategies

Study Objectives

Problem statement——Exploring the Relationship FPG Levels and Hospitalization

- Evaluating how glucose abnormalities affect hospitalization days, aiming to quantify their impact on healthcare resources utilization

- Standard multicenter data in OHDSI to understand the relationship between plasma glucose levels and hospitalization days. This uses OHDSI's multicenter data to study how plasma glucose levels relate to the length of hospital stays, identifying trends for better care planning

Study Design

The study population will be defined as:

1. Patients aged ≥18 years, of any gender.
2. A documented fasting plasma glucose (FPG) measurement during inpatient stay, including a valid numerical value and timestamp, expressed in mg/dL or mmol/L, and with FPG ≠ 0.

Patients must also have complete hospitalization records with both admission and discharge dates.

Additionally, patients must also meet one or both of the following criteria:

- Admission to a specified clinical department (specialty).
- More than one hospitalization episode.

Exclusions:

Incomplete hospitalization records or hospitalization dates falling outside the past 5 years.

Preliminary Analysis to Determine Feasibility

	Analytic use case	Type	Structure
1. FPG only	**Clinical characterization**	**Disease Natural History**	Amongst patients **with at least one fasting plasma glucose (FPG) measurement within a fixed continuous observation period starting from the event**, what are the patient's characteristics from their medical history?
2. Glucose measurement in hospital	**Clinical characterization**	**Disease Natural History**	Amongst patients **who experience an inpatient or ER visit accompanied by glucose measurement during the same clinical encounter**, what are the patient's characteristics from their medical history?

CohortDiagnostics

- URL of CohortDiagnostics package:
  - https://github.com/ohdsi-studies/2025APACStudy-Fudan/tree/master/CohortDiagnostics
- For troubleshooting:
  - Open an issue at https://github.com/ohdsi-studies/2025APACStudy-Fudan/issues
- Send your results to:
  - Changran Wang crwang@fudan.edu.cn, Jiaqi Liu liu.jiaqi@zs-hospital.sh.cn
  - OHDSI APAC apacsymposium@ohdsi.org
- Join us at our study Teams channel: 2025 APAC Study 1 - Fudan

2025 OHDSI APAC Symposium

December 6-7 • Shanghai Jiao Tong University, China

Event Landing Page

Now Available!!

Agenda *(Tentative)*

Day 1 (Dec 6)

Tutorial Sessions

- Introduction of OHDSI/OMOP
- OMOP CDM and Vocabulary
- ETL
- Analytics

Day 2 (Dec 7)

Updates & Studies

- OHDSI APAC and Regional Chapter updates
- 2025 APAC Studies: Overviews and Results
- Real-world Data Developments in China

Discussions & Presentations

- Cross-community Panel Discussion
- Collaborator Showcase: Poster Presentations and Lightning Talks

Collaborators Showcase

2025 OHDSI APAC Collaborator Showcase Brief Report Submission Form

Thank you for your interest in the 2025 OHDSI APAC Collaborator Showcase! We are delighted that you are considering joining our research community and presenting your work at this year's symposium. The 2025 OHDSI APAC Symposium will be held in person **December 6-7** at the Shanghai Jiao Tong University in Shanghai, China.

Please take a few minutes to fill out this submission form to help the OHDSI APAC Scientific Review Committee better understand your work. The deadline to submit your brief report is **September 7**. You will receive a confirmation email of your responses upon completion. If the committee has selected your work to be presented at this year's symposium, you will be notified via email by **October 17**.

Should you need to change your responses to any of the questions on this form, please click on the "Edit response" button in the confirmation email you received. Should you need to revise your brief report, please email apacsymposium@ohdsi.org. Your submission will be removed, and you will need to submit again with the revised PDF.

Aug 4 – Sep 7

Abstract Submission

Sep 8 – Oct 6
Oct 7-9 APAC WG workshop

Review by Scientific Committee

Oct 17

Notification of acceptance

Dec 6-7

Symposium

Regional Chapter Mid-year Updates

Malaysia – Steven Young

Updates from Malaysia

1. **MyGenom** has been introduced to the OMOP CDM framework and will begin incorporating the learnings into their genomic and health data initiatives, aligning with global data interoperability standards. Intention is to expand its capabilities into the area of pharmacogenomic.
   - Will be reaching out to the community / team with these experience.

What is the MyGenom Project?

The MyGenom Project is an effort to understand the genetic makeup of Malaysians. It's our country's first large-scale study of population genomics, which means sequencing the DNA of Malaysian citizens from different races and ethnicities. In Phase I, 2,400 genomes will be sequenced from healthy Malaysian individuals, with the goal of reaching 10,000 genomes in Phase II. Over time, this project will create a Malaysia reference genomic dataset that reflects the diversity of our population and will be the key to bringing Precision Medicine to Malaysia.

Updates from Malaysia (continue)

2. **National Institutes of Health (NIH)** is planning to establish its **Trusted Research Environment (TRE)** in **Q1/Q2 2026**. This TRE will model after OMOP CDM principles, emulating the success of TRUST framework, and NIH is currently applying for funding to support the project.
   - Will be reaching out to the community / team with these experience.

3. **Ministry of Health (MOH)** will be embarking on its **first OMOP CDM project** in **Q1 2026**, paving the way for broader adoption of standardized healthcare data models across the public healthcare sector. The first phase will start off with data from public clinics and followed by hospitals. These initiatives is being done in parallel to their nationwide effort in digitalizing clinics across the country.

Deep Dive into OHDSI India's Developments

Swetha J.

INDIA

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI India Chapter

From Raw Data to Reliable Evidence

Version No: 1.0 , Date: 20Aug2025 | © Copyright OHDSI India Chapter

Agenda

Breast Cancer Study Overview

FHIR Work Group Initiative

The CVD Patient Registry

Digital Health Centre of Excellence

Our Team

Dr.Swetha Kiranmayi Jakkuva, Service Leader Research enabling services, GVW

Khansa Fathima – Assistant Professor at JSSAHER

Sai Dileep – Data scientist, GVW

Shreema – Data Engineer, GVW

Sanjay – Healthcare Data analyst, GVW

Sai Pattabhiram – Software Engineer, GVW

Vipina Keloth - Associate Research Scientist at Yale Biomedical Informatics and Data Science

Breast Cancer Study

Problem Statement:

PDF discharge summaries from 2014 to 2022 at JSS Mysore University containing unstructured breast cancer data. This unstructured format presents challenges in accessing, standardizing, and analyzing data for research and clinical insights.

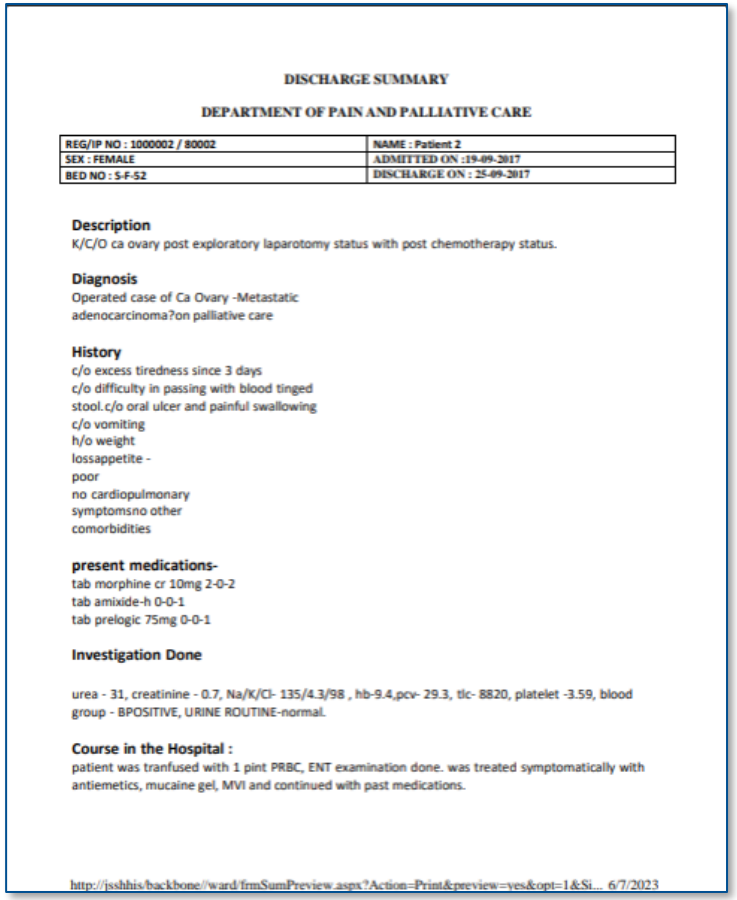
Objectives:

Extract and structure clinically relevant data from the PDF discharge summaries using NLP techniques.

Standardize the extracted data to the OMOP Common Data Model (CDM) for interoperability and analysis.

Enable efficient data utilization for research and real-world evidence (RWE) generation in breast cancer care.

Know our Data



DISCHARGE SUMMARY

DEPARTMENT OF PAIN AND PALLIATIVE CARE

REG/IP NO : 1000002 / 80002	NAME : Patient 2
SEX : FEMALE	ADMITTED ON :19-09-2017
BED NO : S-F-52	DISCHARGE ON : 25-09-2017

Description
K/C/O ca ovary post exploratory laparotomy status with post chemotherapy status.

Diagnosis
Operated case of Ca Ovary -Metastatic
adenocarcinoma?on palliative care

History
c/o excess tiredness since 3 days
c/o difficulty in passing with blood tinged
stool.c/o oral ulcer and painful swallowing
c/o vomiting
h/o weight
lossappetite -
poor
no cardiopulmonary
symptomsno other
comorbidities

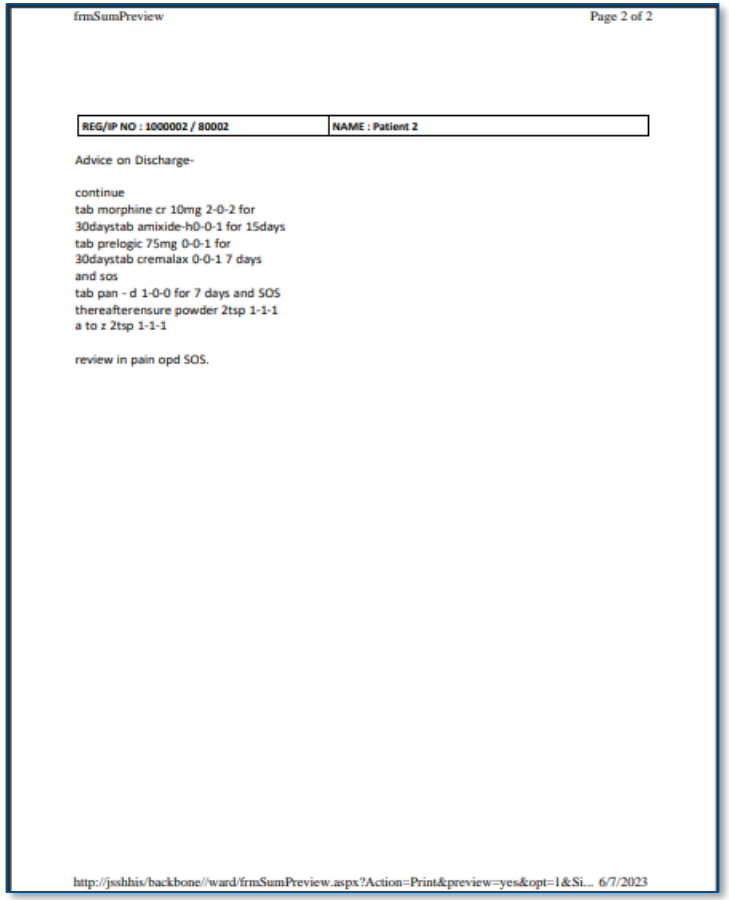
present medications-
tab morphine cr 10mg 2-0-2
tab amixide-h 0-0-1
tab prelogic 75mg 0-0-1

Investigation Done

urea - 31, creatinine - 0.7, Na/K/Cl- 135/4.3/98 , hb-9.4,pcv- 29.3, tlc- 8820, platelet -3.59, blood group - BPOSITIVE, URINE ROUTINE-normal.

Course in the Hospital :
patient was tranfused with 1 pint PRBC, ENT examination done. was treated symptomatically with antiemetics, mucaine gel, MVI and continued with past medications.

http://jsshhis/backbone//ward/frmSumPreview.aspx?Action=Print&preview=yes&opt=1&Si... 6/7/2023



frmSumPreview Page 2 of 2

REG/IP NO : 1000002 / 80002	NAME : Patient 2

Advice on Discharge-

continue
tab morphine cr 10mg 2-0-2 for
30daystab amixide-h0-0-1 for 15days
tab prelogic 75mg 0-0-1 for
30daystab cremalax 0-0-1 7 days
and sos
tab pan - d 1-0-0 for 7 days and SOS
thereafterensure powder 2tsp 1-1-1
a to z 2tsp 1-1-1

review in pain opd SOS.

http://jsshhis/backbone//ward/frmSumPreview.aspx?Action=Print&preview=yes&opt=1&Si... 6/7/2023

- Demographic
- Medical history
- Labs / Diagnostics
- Procedures
- Drugs

Process Overview

- PDF Deidentification to remove PHI
- Conversion of PDF to text
- Annotating the clinical notes
- Developing NLP models - NER, RE, & CN
- Evaluating Model Performance Precision, Recall, F1 score, Accuracy
- OMOP CDM Data Conversion

Process Overview

- PDF Deidentification to remove PHI
- Conversion of PDF to text
- Annotating the clinical notes
- Developing NLP models - NER, RE, & CN
- Evaluating Model Performance Precision, Recall, F1 score, Accuracy
- OMOP CDM Data Conversion

Deidentification

Id : 2fde18bf55ee24c03e7024a5b0b2721055739d02196dc7b7480ed5214ecaa8c5

DISCHARGE SUMMARY

DEPARTMENT OF MEDICINE

REG/IP NO	: [redacted]	NAME	: [redacted]
AGE	: 30 Years	ADMITTED ON	: 23-11-2014
SEX	: FEMALE	DISCHARGED ON	: 27-11-2014
BED NO	: GM-F-01	UNIT HEAD	: [redacted] M MBBS,MD,FACP (USA),FRCP
ADDRESS	[redacted]		
	MYSURU		

GENERAL MEDICINE

FINAL DIAGNOSIS: Ca Breast with ?Secondaries to the Brain.

PRESENTING SYMPTOMS: Headache- 10 days
Weakness of left side of body-8 days back lasted for 10 min.
Fever- 2days
Diplopia -2 days

PAST HISTORY: K/C/O of Ca Breast 1 year back for which she got operated in month of Aug 2013 followed by 2 cycles of chemotherapy.

EXAMINATION:

GPE- PR : 90bpm BP : 118/70 mmhg
CVS- NAD
RS - NAD
PA- NAD
CNS-Conscious and oriented.HMF-normal
Power Rt UL and LL-5/5
Lt UL and LL-4/5
Plantar Rt-flexor lt-extensor

INVESTIGATIONS: Enclosed

TREATMENT GIVEN: Inj Mannitol 100ml i.v 1-1-1
Tab Strocit-P 1-0-1
Tab Ultracet 1-0-1
Tab Rantac 150 mg 1-0-1
Tab Eptoin 100mg 1-0-2
Inj Dexona 8mg stat and 4mg 1-1-1

A specialized deidentification tool is used to mask the PHI from PDF discharge summaries, ensuring patient privacy and compliance with data security standards.

Process Overview

- PDF Deidentification to remove PHI
- Conversion of PDF to text
- **Annotating the clinical notes**
- Developing NLP models - NER, RE, & CN
- Evaluating Model Performance Precision, Recall, F1 score, Accuracy
- OMOP CDM Data Conversion

Develop Annotation Guidelines

Defined the main entities and modifiers based on the purpose of annotation.

Entity Name	Definition
Problem	The abnormal condition that happens physically or mentally to a patient.
Procedure	A medical process or method used to perform a test or treatment of a problem.
Drug	Generic or brand name of a single medication or a collective name of a group of medication.
Test	A medical procedure performed (i) to detect or diagnose a problem, (ii) to monitor diseases, disease processes, and susceptibility, or (iii) to determine a course of treatment.

Modifier Name	Definition
Severity	The degree of intensity of a clinical condition
Conditional	A phrase that indicates the problems existing in a certain situation
Uncertain	A measure of doubt related to a problem
Bodyloc	The location on the body where the observation is present
Course	The development or alteration of a problem
Temporal	Calendar date, time, or duration related to a problem
Negation	The phrase that indicates the absence of an entity
Problem Status	Indicates the temporal status of the problem, distinguishing between current and past conditions.

Develop Annotation Guidelines

Basic annotation rules

Example - Annotate demographic data such as gender (male/female) and age information

Specific guidelines for each main entity

Problem - Definition: The phrases containing the observations made by the patient or clinician about the patient's body or mind as abnormal or caused by a disease are considered as a "Problem".

Special cases and document as you annotate

Chemotherapy should be annotated as a procedure entity, while the specific drugs used in the chemotherapy should be annotated as drug entities.

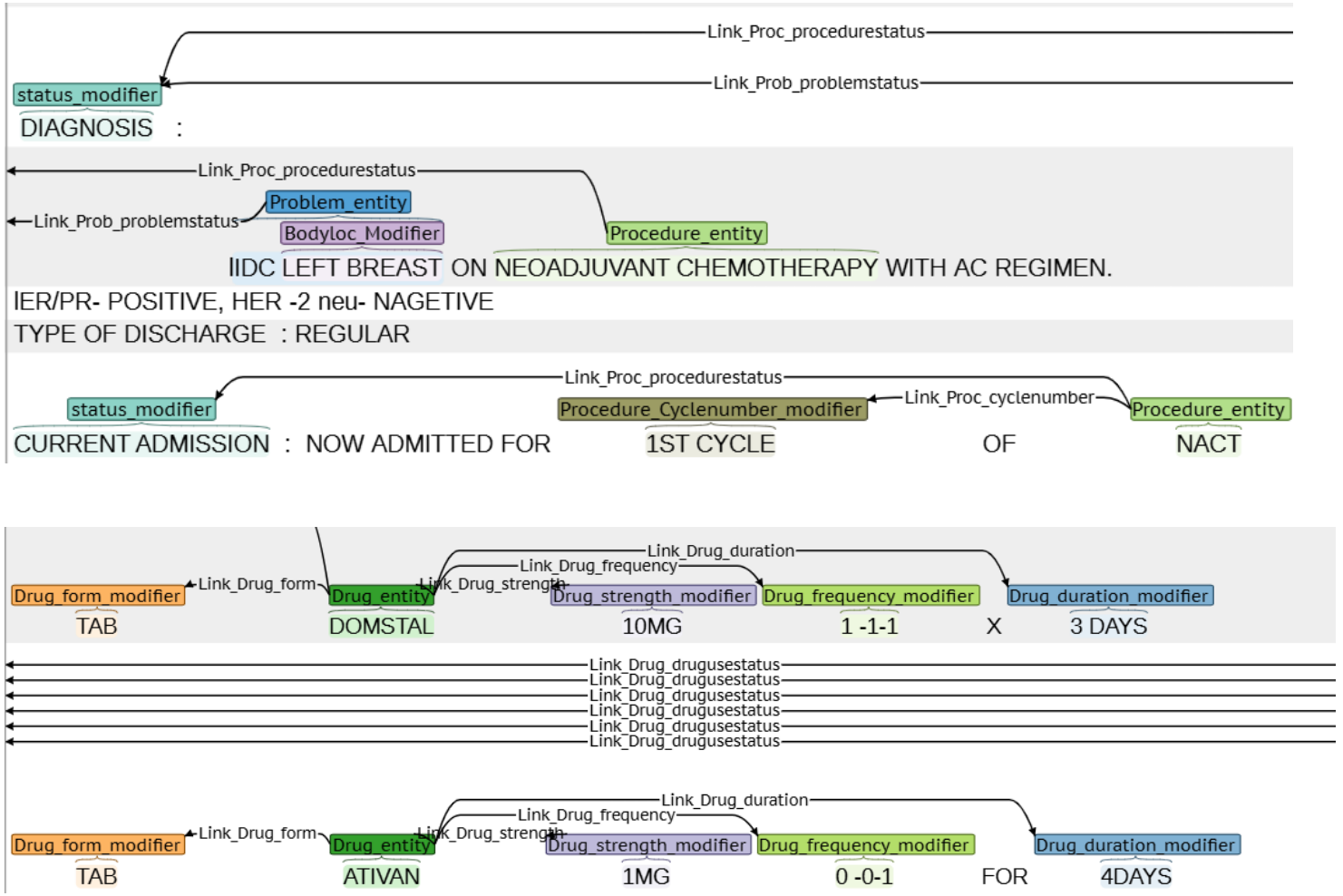
The modifiers

When an ENTITY is described by a MODIFIER, the MODIFIER provides additional information about the ENTITY

Span relationship between modifiers and entities

After annotating the entity, attribute the relationships between the associated modifiers.

Annotations



Annotation Schema

You need to prepare the annotation schema.

Schema: **Breastcancer_Annotation**

- **27** Entity Tags

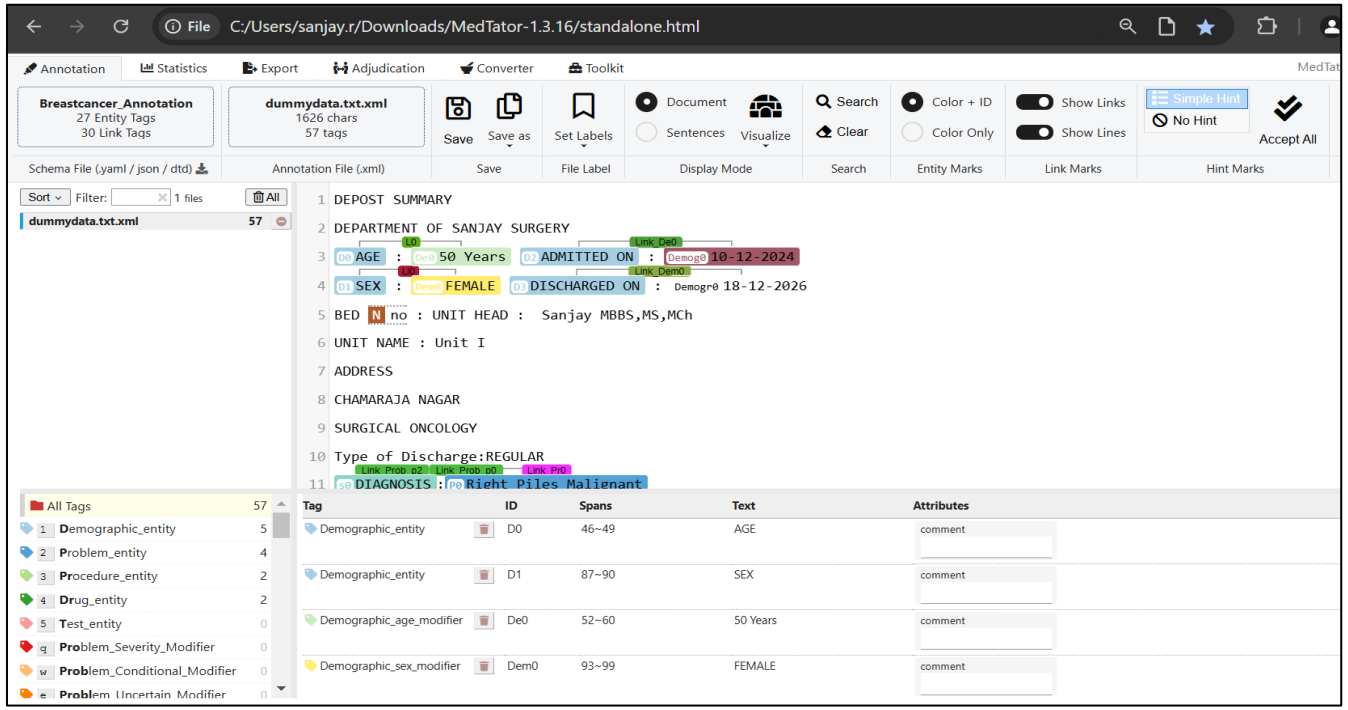
Demographic_entity, Problem_entity, Procedure_entity, Drug_entity, Test_entity, Problem_Severity_Modifier, Problem_Conditional_Modifier, Problem_Uncertain_Modifier, Bodyloc_Modifier, Problem_Course_modifier, Temporal_modifier, Negation_modifier, status_modifier, Procedure_Cyclenumber_modifier, Drug_strength_modifier, Drug_dosage_modifier, Drug_duration_modifier, Drug_form_modifier, Drug_frequency_modifier, Drug_route_modifier, Test_Labvalue_modifier, Test_Referencerange_modifier, Demographic_age_modifier, Demographic_sex_modifier, Demographic_uniqueid_modifier, Demographic_Admissiondate, Demographic_Dischargedate

- **30** Link Tags Link_Demo_age, Link_Demo_sex, Link_Demo_uniqueID, Link_Prob_severity

Annotation Tool & Methodology

MedTator

GitHub /OHNLP/MedTator



Results:

Inter-Annotator Agreement (IAA) was measured across different batches of annotated files. The results are as follows:

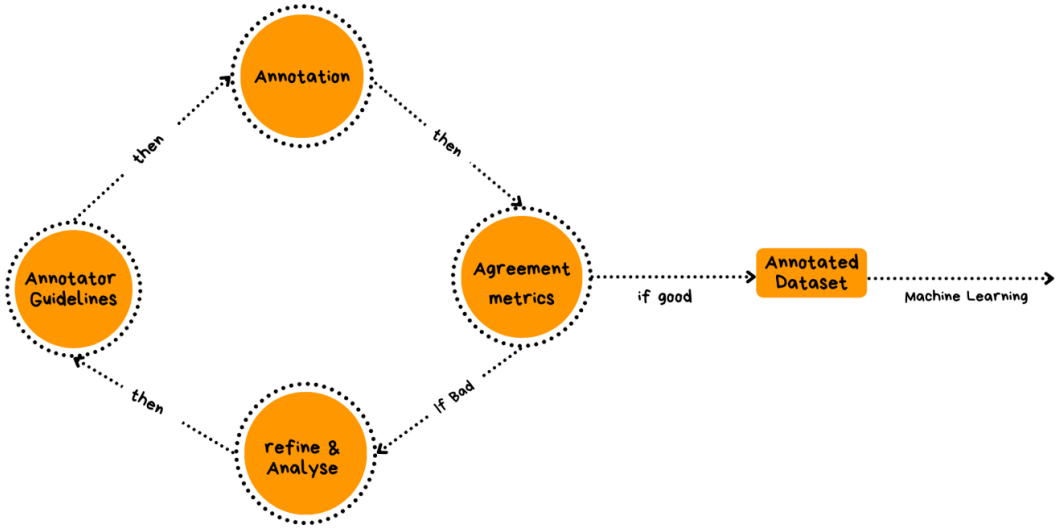
IAA (1): Annotator A vs. Annotator B (Batch of 25 files) - **94.71%**
IAA (2): Annotator A vs. Annotator C (Batch of 75 files) - **96.12%**
IAA (3): Annotator A vs. Annotator D (Batch of 75 files) - **93.10%**
IAA (4): Annotator A vs. Annotator D (Batch of 125 files) - **94.32%**

Error Analysis :
Error analysis was conducted by Annotator A to maintain quality control and ensure accurate annotations.

Inter-annotator Agreement (Adjudication)

MedTator

- **Cohen's Kappa**
- **Krippendorff's alpha**



$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

	tag_name	F1	precision	recall	TP	FP	FN
2	Overall	0.8866	0.8337	0.9468	2436	486	137
3	Demographic_ent	0.9540	0.9120	1.0000	114	11	0
4	Problem_entity	0.6405	0.5024	0.8833	106	105	14
5	Procedure_entity	0.7547	0.6557	0.8889	80	42	10
6	Drug_entity	0.8931	0.8224	0.9771	213	46	5
7	Test_entity	0.9507	0.9450	0.9564	395	23	18
8	Problem_Severity	1.0000	1.0000	1.0000	1	0	0

Process Overview

- PDF Deidentification to remove PHI
- Conversion of PDF to text
- Annotating the clinical notes
- **Developing NLP models - NER, RE, & CN**
- Evaluating Model Performance Precision, Recall, F1 score, Accuracy
- OMOP CDM Data Conversion

Data extraction from the discharge summaries

Named entity recognition

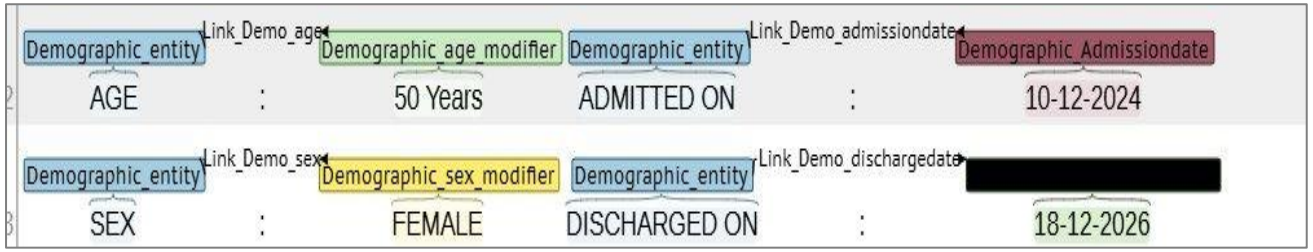
Identifies 27 modifiers across demographics, problems, procedures, tests, and drugs, including sub-modifiers.

Relation extraction

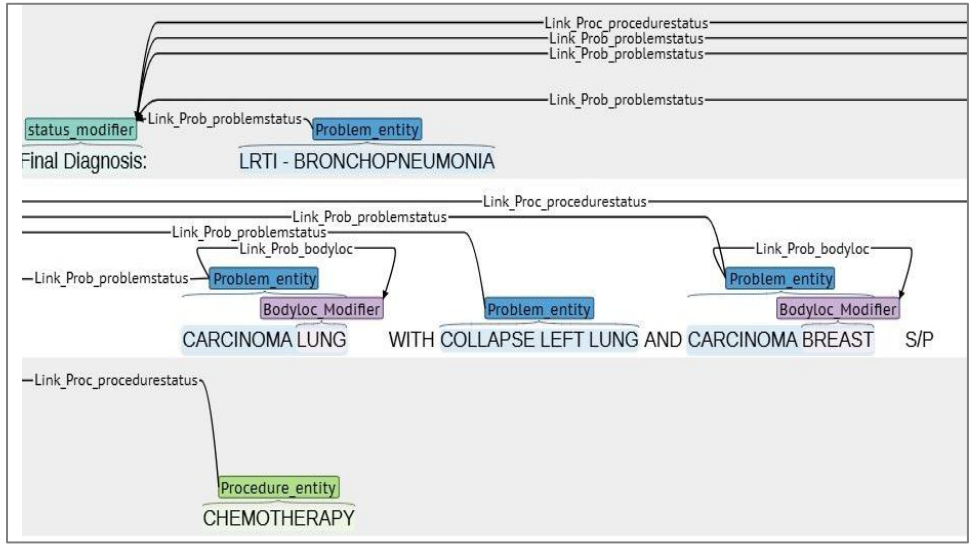
Maps relationships like historical vs. current problems and determines severity.

Concept Normalization and Vocabulary Mapping

Normalizes and aligns entities to OMOP Concept IDs for standardization



TRAIN:VALIDATE:TEST – 70:10:20

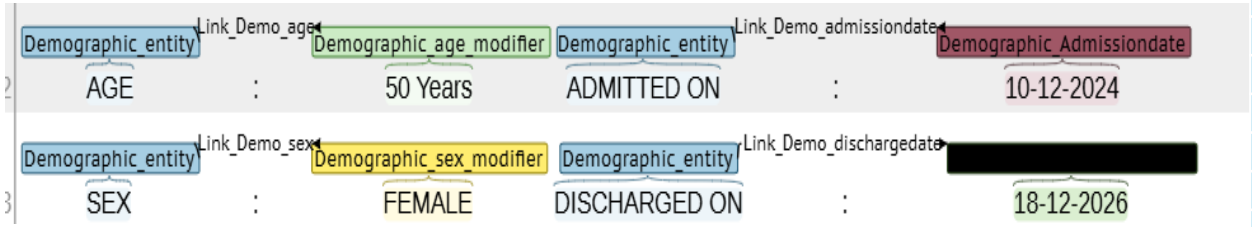


Named Entity Recognition

BERT Model

Identifying the boundaries of entities

Identifying the semantic categories of entities



BIO (B – Beginning I – Inside O – Outside) Tagging

Token	Label
AGE	B-Demographic_entity
:	O
50	B-Demographic_age_modifier
Years	I-Demographic_age_modifier
ADMITTED	B-Demographic_entity
ON	I-Demographic_entity
:	O
10-12-2024	B-Demographic_Admissiondate
SEX	B-Demographic_entity
:	O
FEMALE	B-Demographic_sex_modifier
DISCHARGED	B-Demographic_entity
ON	I-Demographic_entity
:	O
18-12-2026	B-Demographic_Dischargedate

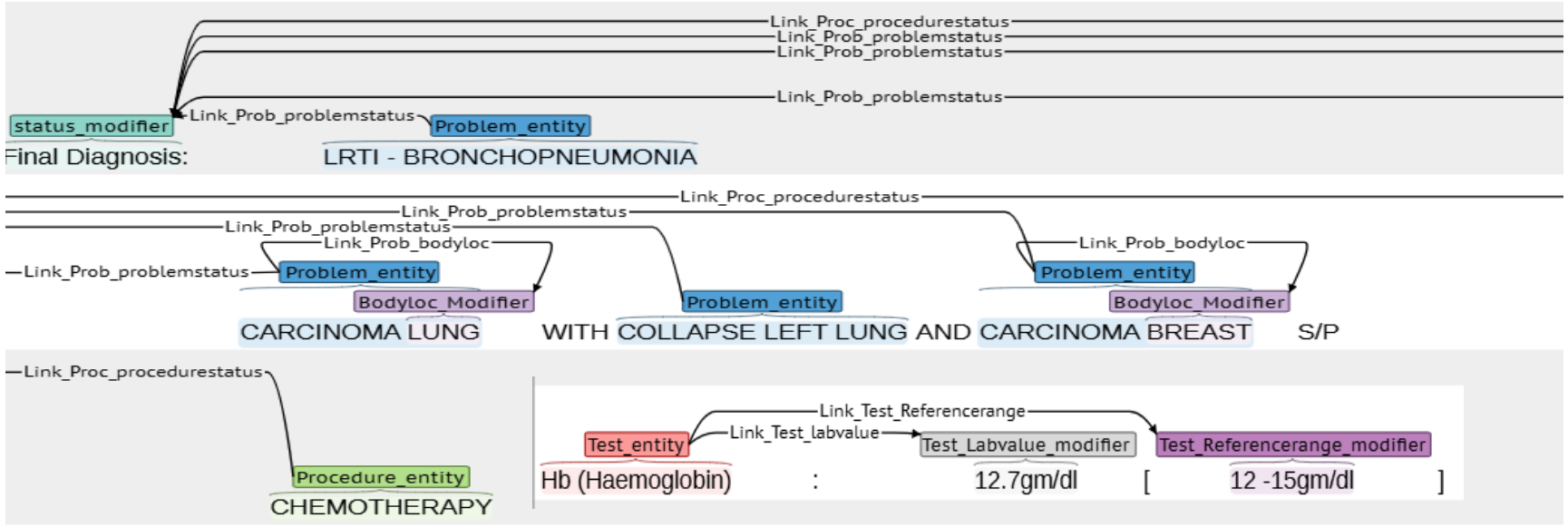
Evaluation

Text	Gold Annotation	Annotation	Result
The patient was diagnosed with **breast cancer**	breast → B-Problem cancer → I-Problem	breast → B-Problem cancer → I-Problem	✅ **Exact Match** (same entity type + same boundaries)
The patient was diagnosed with **breast cancer**	breast → B-Problem cancer → I-Problem	cancer → B-Diagnosis	🟡 **Relaxed Match** (entity type matches, boundaries overlap but not identical)
The patient was diagnosed with **breast cancer**	breast → B-Problem cancer → I-Problem	breast → B-BodyPart	❌ **Not a Match** (entity type mismatch, even though text overlaps)

Relationship Extraction

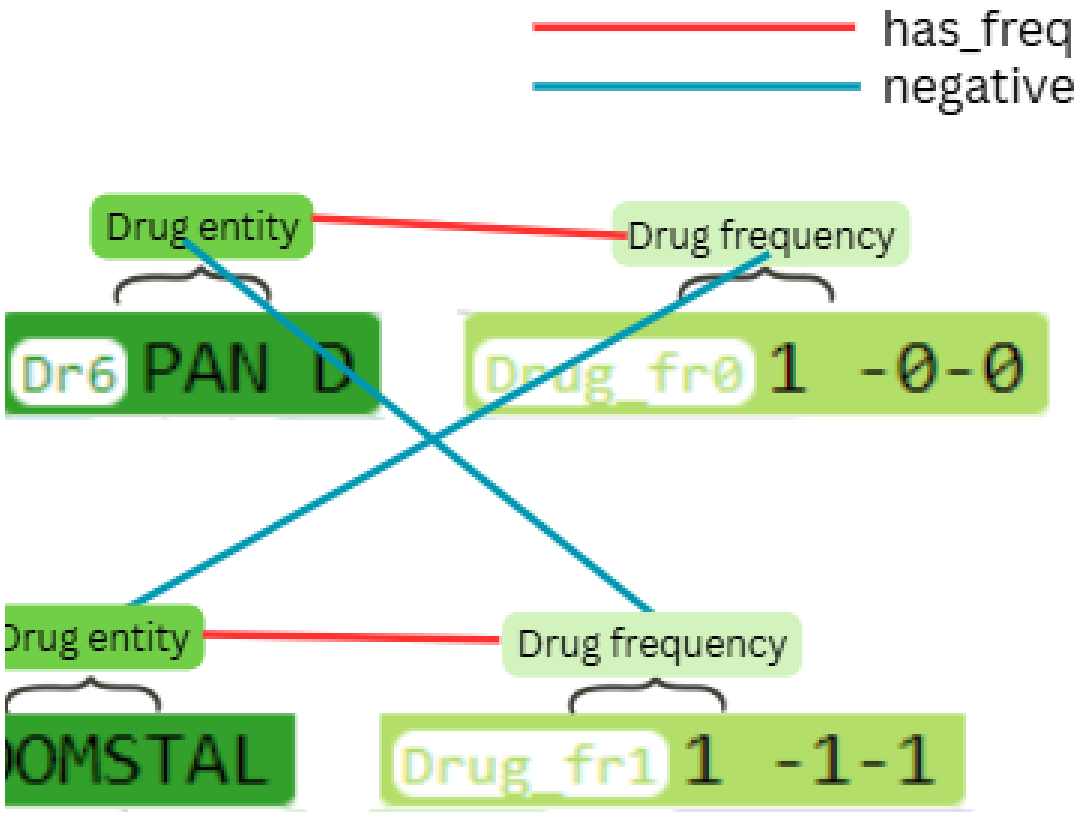
BERT Model

- Identify the semantic relations between entities.
- Determine the direction and context of the relation



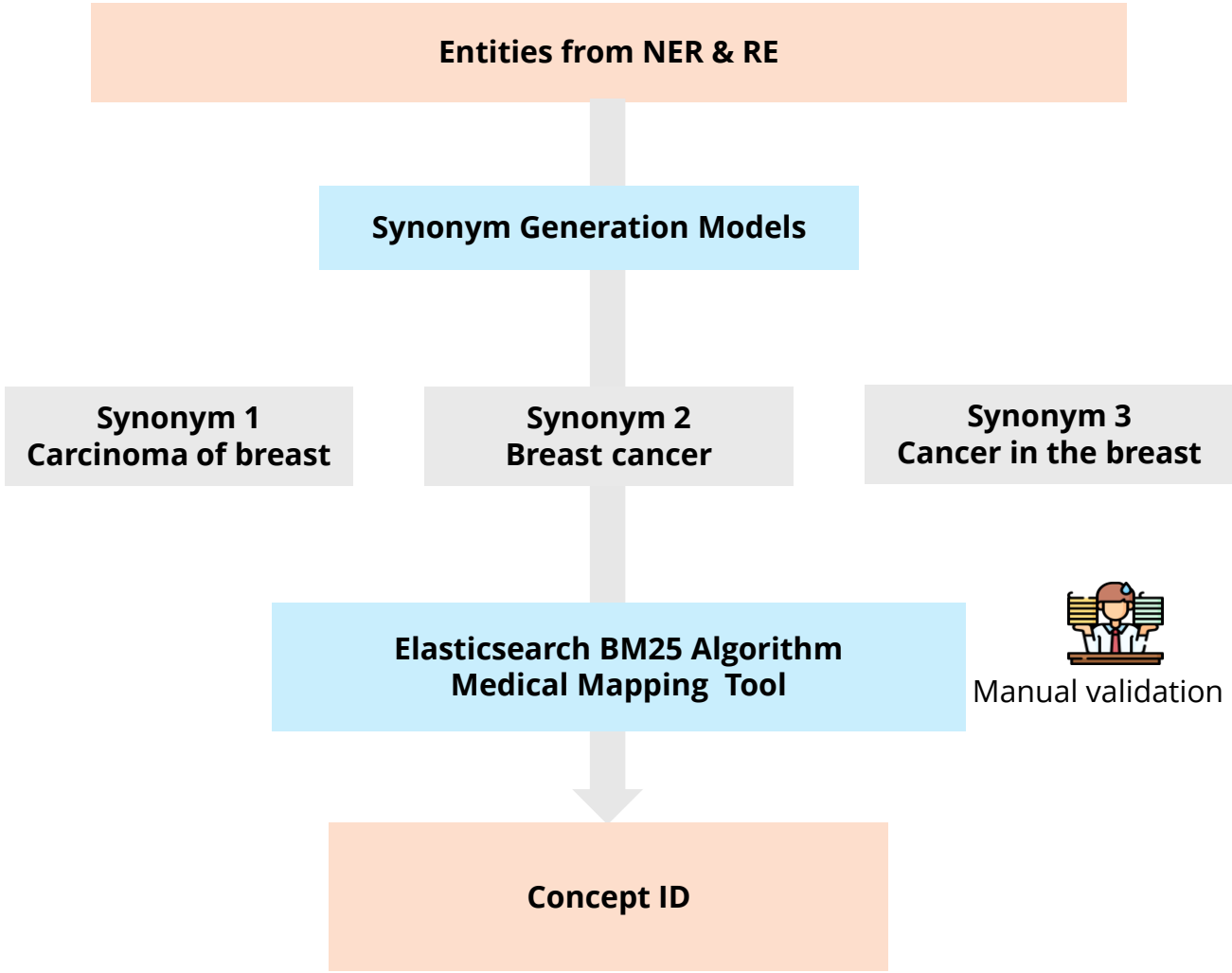
Entity	Attribute Value	Relationship
Carcinoma	Breast	Has_attribute
LRTI -Bronchopneumonia	Final Diagnosis	Has_attribute
Hb(Haemoglobin)	12.7gm/dl	Has_attribute

Evaluation



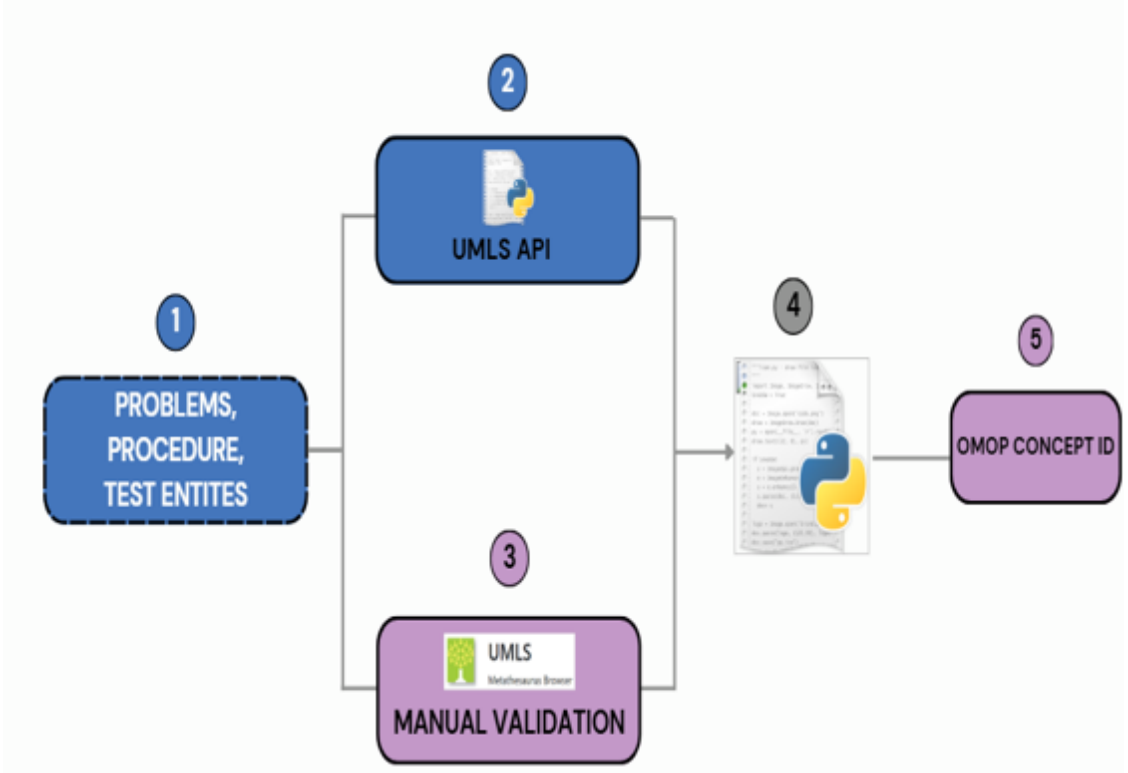
- **Sentence classification of 2 classes:**
- **[Has_attribute, Negative]**

NLP Step 3 - Concept Normalization & Vocabulary Mapping

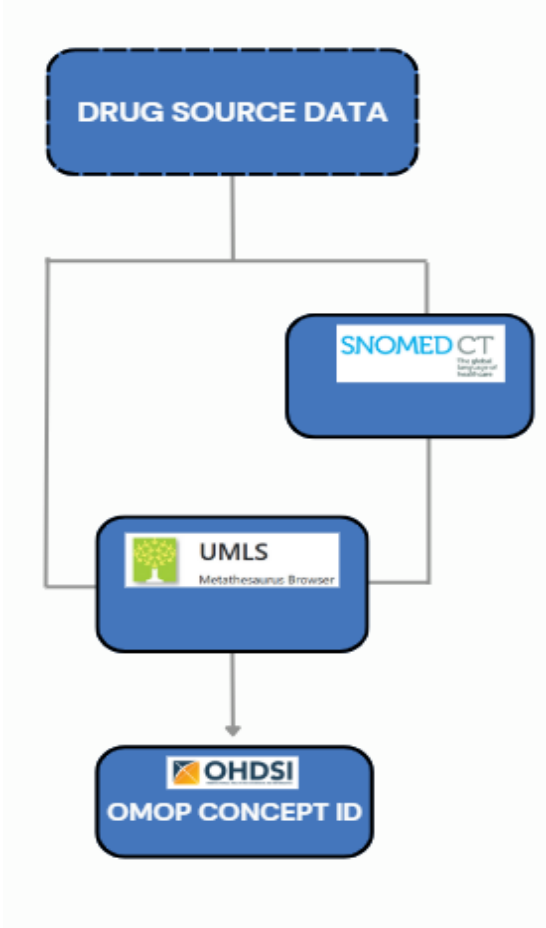


NLP Step 3 - Concept Normalization & Vocabulary Mapping

Clinical Text



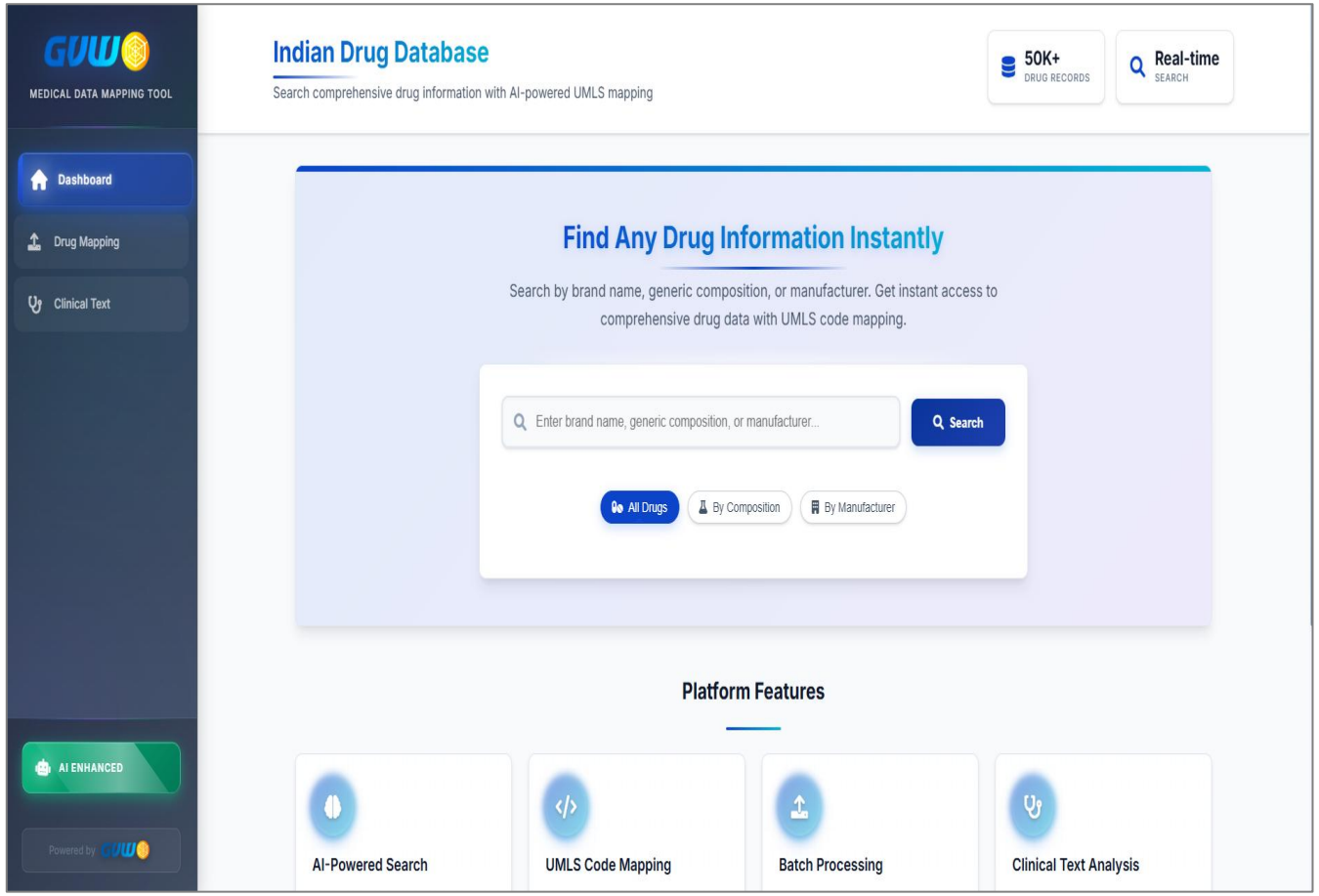
Drugs



Medical Data Mapping Tool

Features –

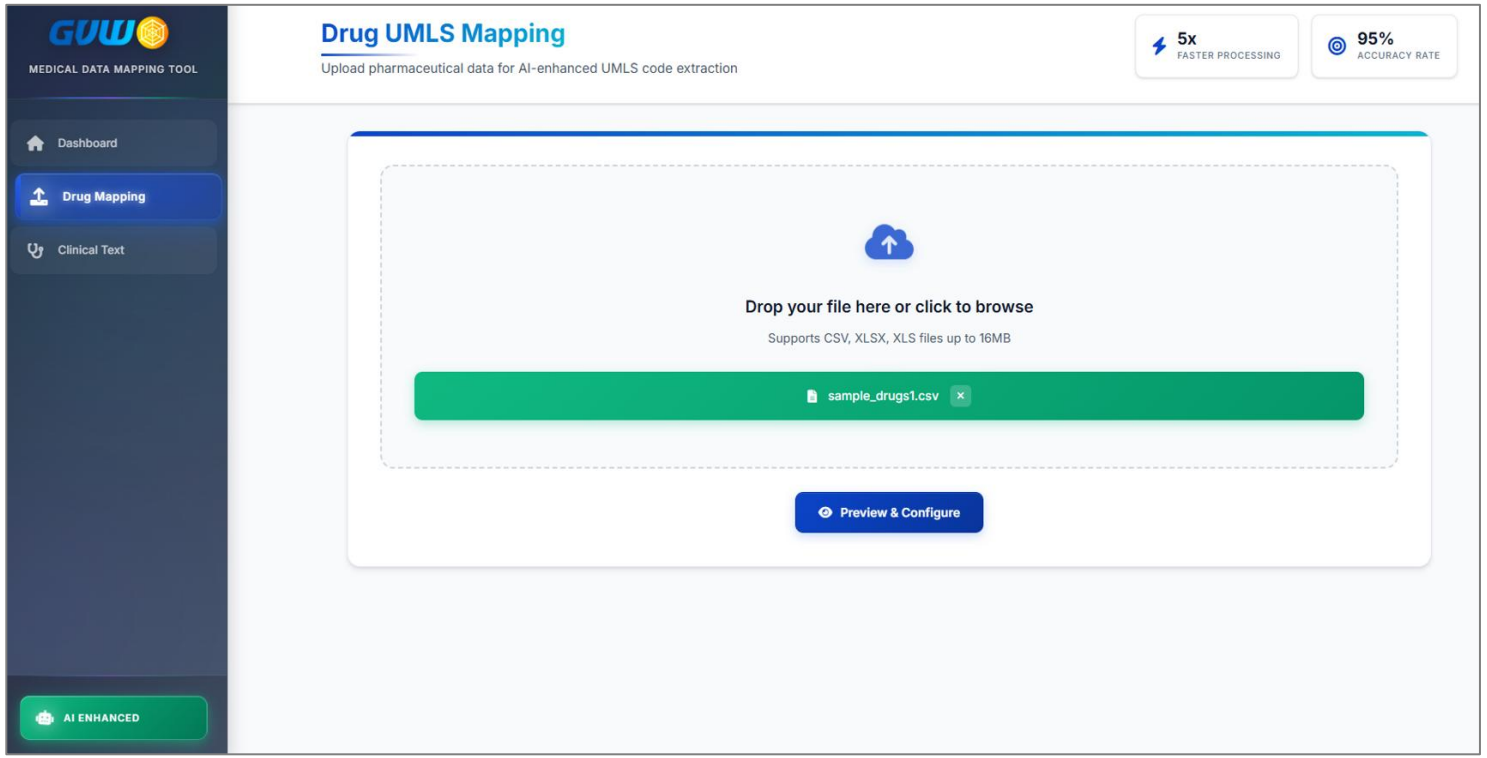
- Maps terms to UMLS, RxNorm, and SNOMED CT.
- High accuracy with Gemini AI.
- Manual review and correction option.
- Auto spell-check and standardization.
- Identifies and standardizes abbreviations
- Maps Indian brand names to generics.
- Scalable for large datasets.



Medical Data Mapping Tool - Input

Input

drug	strength
Pantoprazole	40 mg
Dolo	650 mg
Atorvastatin	40 mg
Ultracet	
Aceclofenac	
Telma	40 mg
Terbinafine	
Cetrizine	10mg
Atorva	20 mg
Metformin	500 mg
Pan	40 mg
Paracetamol	650 mg



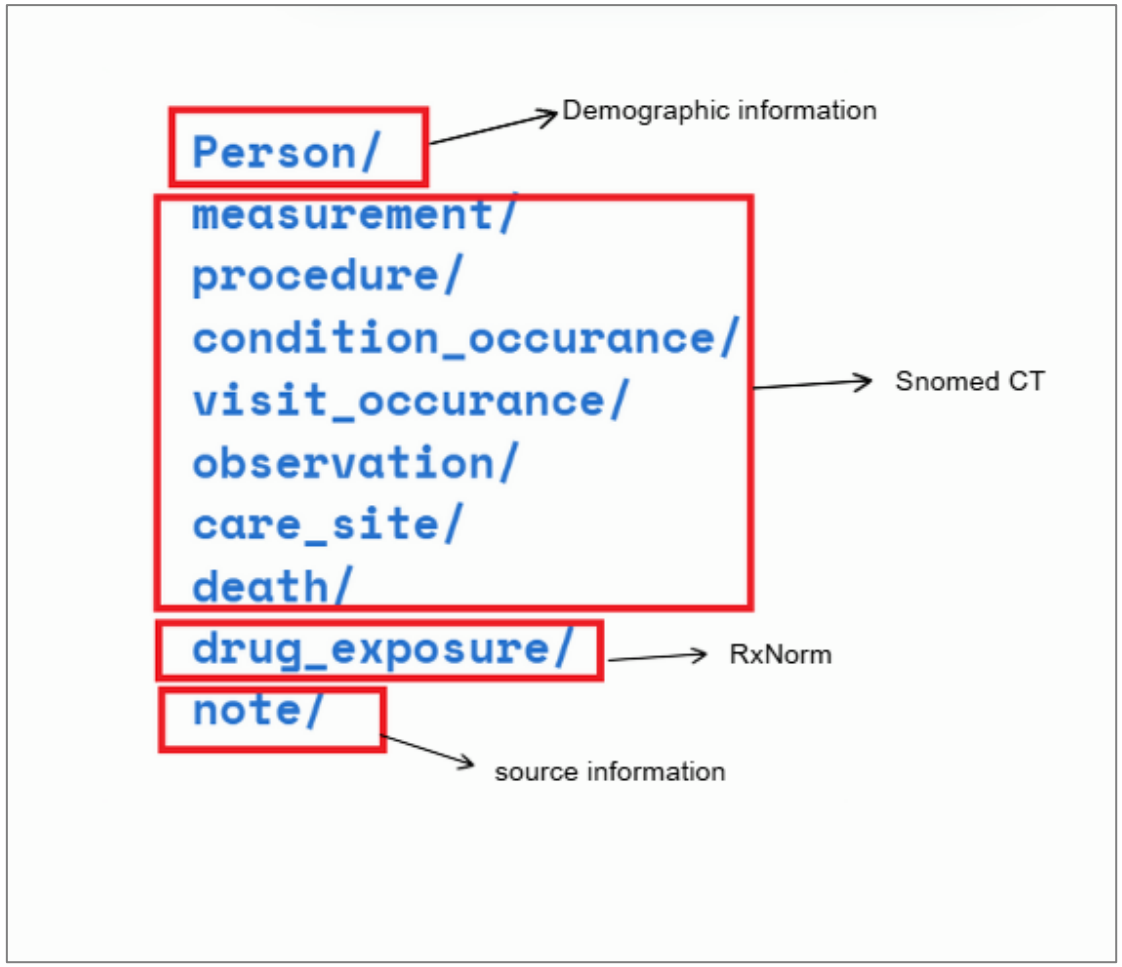
Medical Data Mapping Tool - Output

Term	CUI	Name	Source	Source Code	Semantic Types
pantoprazole 40mg	C1126048	pantoprazole 40 MG	RXNORM	330396	Clinical Drug
Dolo 650mg	C0691095	acetaminophen 650 MG Oral Tablet	RXNORM	198444	Clinical Drug
Atorvastatin 40mg	C1667395	atorvastatin 40 MG	RXNORM	597983	Clinical Drug
Ultracet	C1253063	acetaminophen / tramadol Oral Tablet	RXNORM	378712	Clinical Drug
acelofenac	C0050403	aceclofenac	RXNORM	16689	Organic Chemical|Pharmacologic Subs
telma 40mg	C0990502	telmisartan 40 MG	RXNORM	316764	Clinical Drug
terbinafine	C0076110	terbinafine	RXNORM	37801	Organic Chemical|Pharmacologic Subs
cetrzine	C0055147	cetirizine	RXNORM	20610	Organic Chemical|Pharmacologic Subs
Atorva 20mg	C0286651	atorvastatin 20 mg	RXNORM	83367	Organic Chemical|Pharmacologic Subs
Metfomrin 500mg	C0025598	metformin	RXNORM	6809	Organic Chemical|Pharmacologic Subs
Pan 40mg	C0081876	pantoprazole	RXNORM	40790	Organic Chemical|Pharmacologic Subs
Paracetamol 650	C0691095	acetaminophen 650 MG Oral Tablet	RXNORM	198444	Clinical Drug

Process Overview

- PDF Deidentification to remove PHI
- Conversion of PDF to text
- Annotating the clinical notes
- Developing NLP models - NER, RE, & CN
- Evaluating Model Performance Precision, Recall, F1 score, Accuracy
- **OMOP CDM Data Conversion**

OMOP CDM Table Structure



What We Learned

De-identification	Removed patient-identifiable information from clinical text
Named Entity Recognition	Extracted 5 main Entities : problem, test, medication, procedure, demographic details
Relationship Extraction	25 modifiers extracted
	Examples: temporal, disease, drug strength, dosage, frequency
	Contextualized: past vs current (based on discharge summary date)
Concept Normalization & Vocabulary Mapping	Mapped Indian drugs to OMOP concept IDs
	Built tool to convert extracted text/drugs to OMOP vocabulary

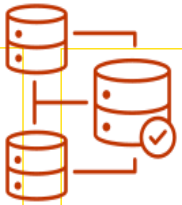
Realities of Research

Evolving data privacy laws limited broader collaboration

Unable to predict infrastructure needs for model development

Resource-intensive, managed by a lean team

Excessive time spent on manual data annotation



Challenges with Indian drug mappings and standardization

Next steps

- Prepare and submit a publication on the study experience.
- Automate discharge summaries into standardized OMOP data.
- Provide an ATLAS overview and training to JSS researchers.
- Explore opportunities for networking and collaborative studies.
- Train and expand medical mapping tool for wider community use.

Preliminary Research Questions

Cohort characterization:

What are the demographic profiles, common comorbidities, procedures, drugs, and tests among inpatients with breast cancer?

Treatment patterns:

Among patients with [target condition], what procedures and discharge drugs are most commonly used, and how do patterns vary by age/sex?

Length of Stay Prediction

Use available features to predict prolonged hospital stays in breast cancer patients.

Prediction / Risk Profiling

Can baseline demographic and clinical features predict whether a patient will complete ≥6 chemotherapy cycles?

FHIR Work Group Initiative

Current team – Kumar Satyam (Associate Director at Providence India, HL7 Chair), Dr.Chandil (Co-founder at SVM hospital), Manish (Founder Yajur Healthcare), Parthiban (VP- Innovation & Growth, GVW), Swetha Kiranmayi (Service leader, GVW)

Objectives

ABDM-FHIR to OMOP Conversion

Enable structured mapping from ABDM-compliant FHIR to OMOP CDM for scalable, interoperable health data analytics

Global FHIR Guideline Adoption

Localize international FHIR standards to align with India's digital health architecture and support OMOP integration.

Advance Capacity Building and Community Engagement

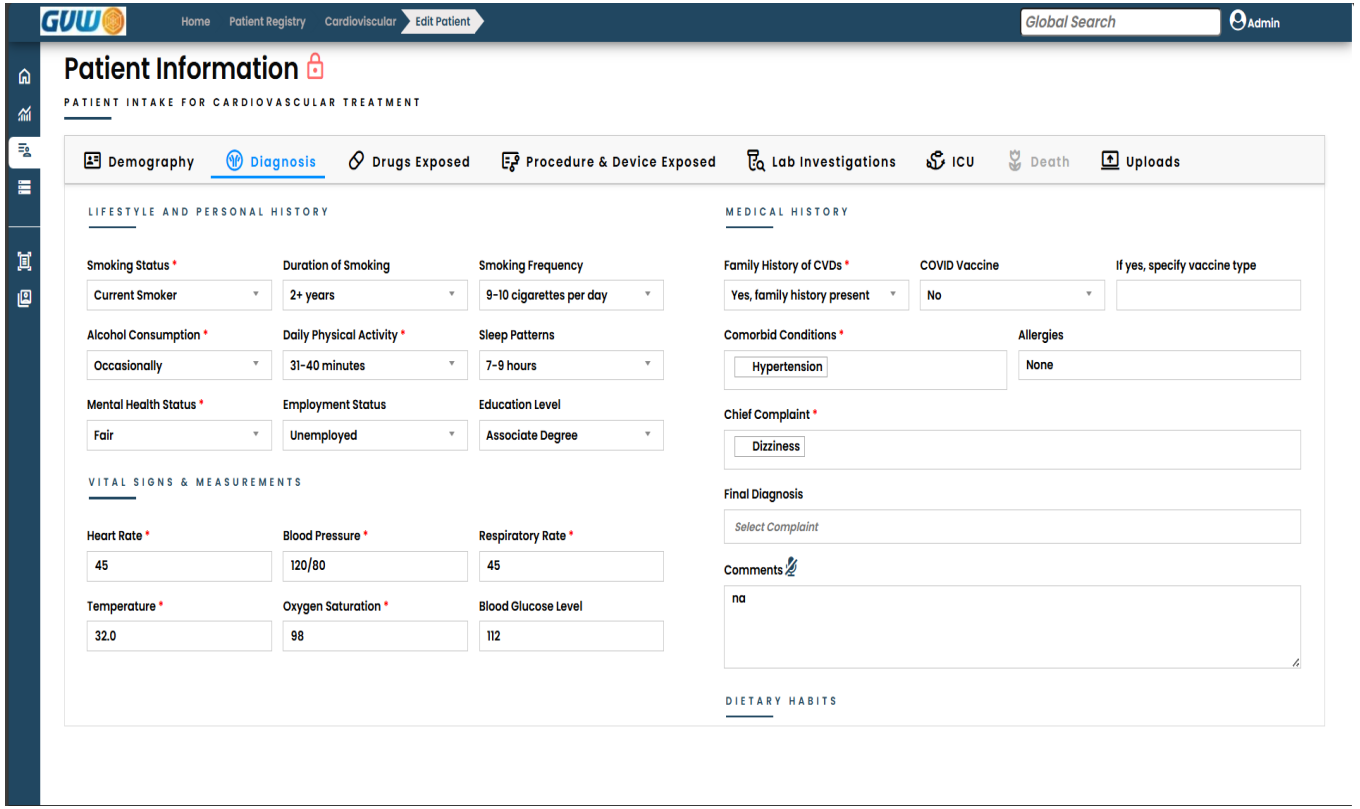
Foster knowledge exchange through workshops, documentation, and open-source contributions

The CVD Patient Registry

Current team – Dr.Vikram Patil (Dy Dean,Research(Clinical), JSS AHER), Dr.Akshay (HOD Critical Care Medicine JSS Hospital), Dr.sunil Kumar (Cardiology specialist at JSS Hospital), Ranjani (Data manager at JSS Hospital), Dr. Swetha Kiranmayi (Service leader, GVW) Sushanth (Senior application engineer, GVW)

Features –

- Enhanced data capture & gap reduction
- OMOP compatibility
- Voice-to-text input
- Predefined fields with medical dictionary integration
- Automated backend coding
- Building capabilities for ABDM-FHIR readiness



Digital Health Centre of Excellence

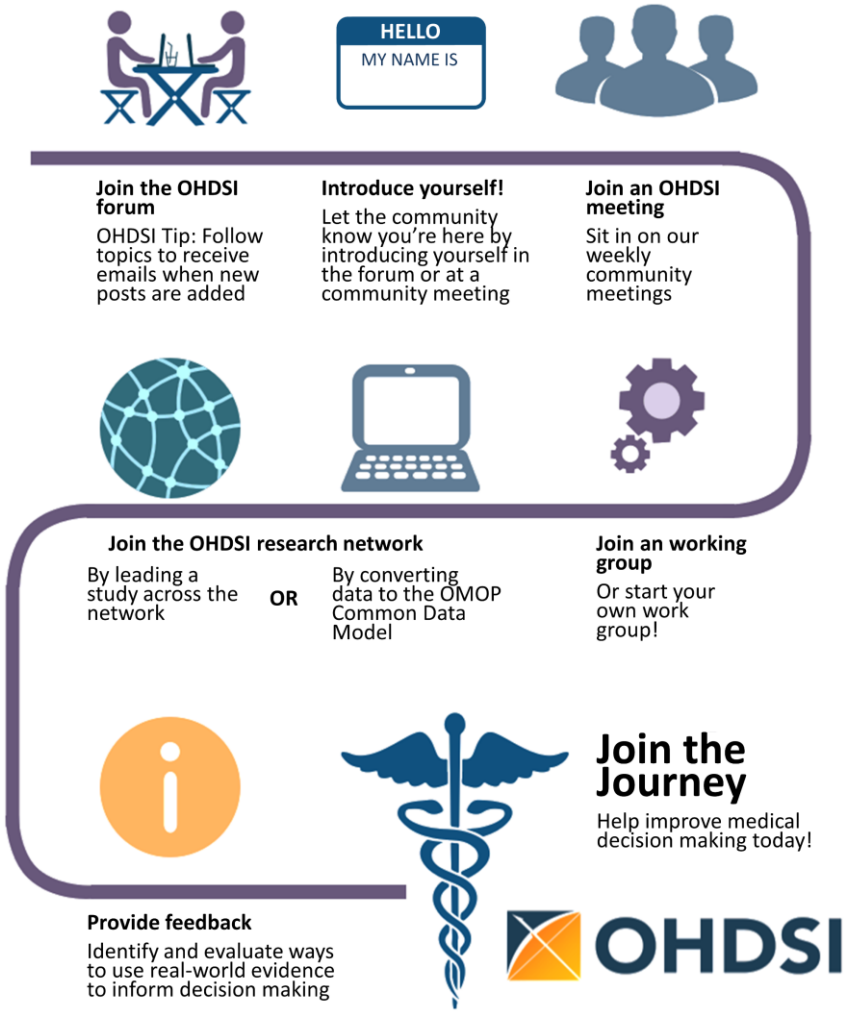
Current team – Rintu Kutum (Group leader, Augmented Health Systems Laboratory. Faculty Fellow of Computer Science. Koita Centre for Digital Health at Ashoka. Mphasis AI & Applied Tech Lab at Ashoka. Trivedi School of Biosciences. Ashoka University), Parthiban (VP-Innovation & Growth, GVW), Swetha Kiranmayi (Service leader, GVW)

Aim – To unlock the potential of large-scale health data in India by enabling real-world data (RWD) generation

Objectives

- Establish a collaborative hub for digital health innovation, standardization, and capacity-building.
- Drive ABDM & FHIR-aligned digital health technologies.
- Engage experts and stakeholders across clinical, technical, and academic domains.
- Use impactful case studies to inform policy and best practices.
- Build an integrated ecosystem combining data, technology, and evidence.
- Promote open science by linking FHIR & OMOP for reproducible research.
- Advance semantic interoperability for Indian health data integration and reuse.
- Secure funding to sustain and scale OHDSI India initiatives.

Join the journey - How to become an OHDSI collaborator



INDIA

Register here: [Be a Collaborator]