



Bakar Computational Health
Sciences Institute

Natural language processing for inferences from electronic health record notes

March 2, 2023

Madhumita Sushil, PhD
Postdoctoral Scholar, Butte lab,
University of California, San Francisco
@madhumitasushil

Real-world data for generating regulatory evidence

Real-world data from EHRs provide valuable evidence to support regulatory decision making

However, the use of real-world data remains challenging:

- Complexity of processing large, temporal data sources
- Complexity of linking different data sources
- Complexity of phenotyping from textual data sources
- Data repurposing may not always be ideal

GUIDANCE DOCUMENT

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Draft Guidance for Industry

Real-world data for generating regulatory evidence

Real-world data from EHRs provide valuable evidence to support regulatory decision making

However, the use of real-world data remains challenging:

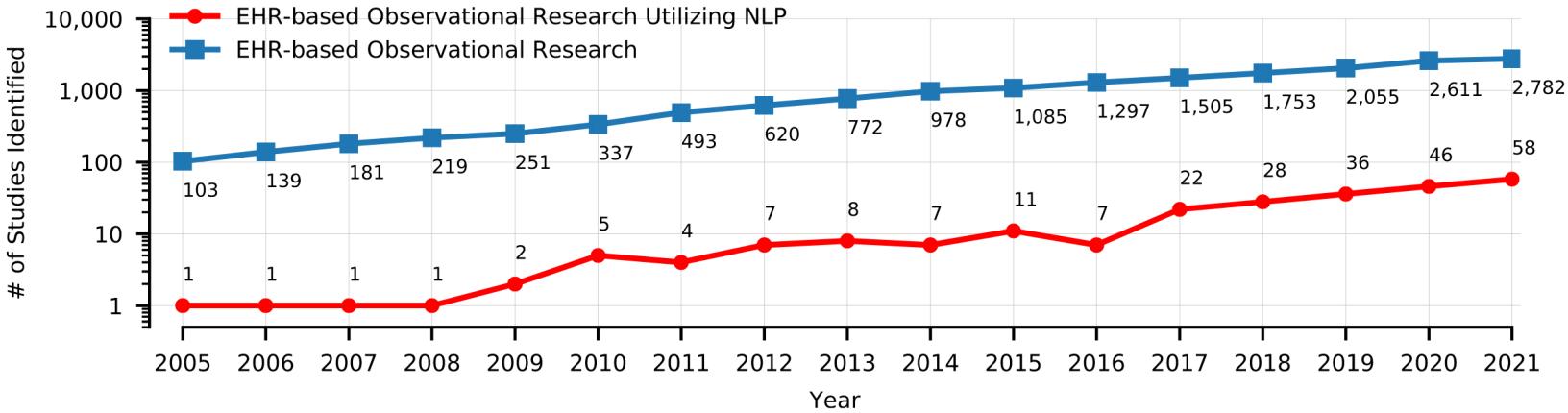
- Complexity of processing large, temporal data sources
- Complexity of linking different data sources
- **Complexity of phenotyping from textual data sources**
- Data repurposing may not always be ideal

GUIDANCE DOCUMENT

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Draft Guidance for Industry

Natural Language Processing is underutilized

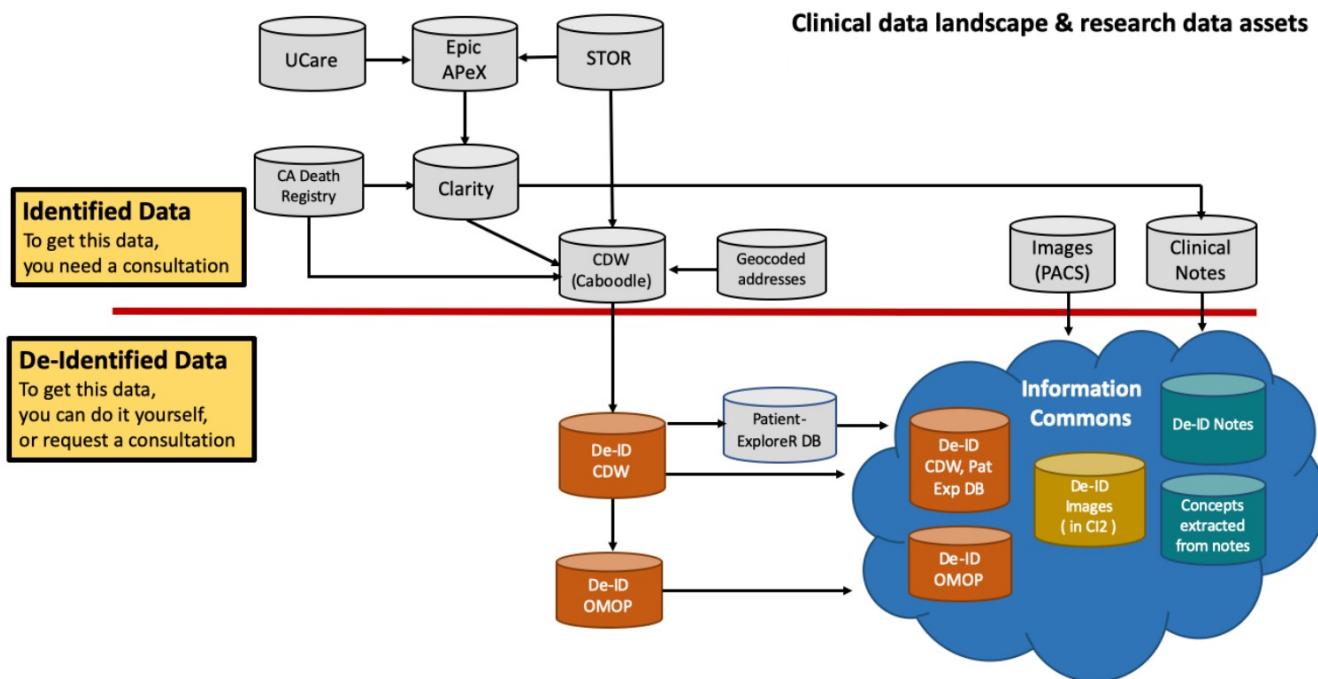


Only 1-2% of all EHR-based observational research use NLP.

Why is NLP underutilized?

- Access to clinical reports
 - Most EHR data is only present in identified formats with PHI, deidentified reports are scarce.
 - Large, publicly-available de-identified clinical reports are scarce. Reduces technical advances from groups without affiliations to medical centers.
 - Clinical text reports are complex, long, and a high-volume temporal data source
 - Both medical as well as technical expertise needed to effectively utilize these reports
 - State-of-the-art tools in generic language processing do not always work for clinical language processing
 - Domain-specific semantics/coded language use: ++ (high), - (absent)
 - Domain-specific abbreviations: US, CD10, BP, Wt, SpO2, prn
 - Ungrammatical structure:
 - » Semi-structured text is sometimes present (e.g. medication reconciliation tables pulled into text reports automatically)
 - » Missing end-of-sentence boundaries (e.g. in reports earlier)
 - » Missing subjects (e.g. *Will refer for EKG*).
 - Extensive manual efforts needed to create supervised annotated datasets and tailor tools to the clinical domain.

Data sources available for research at UCSF



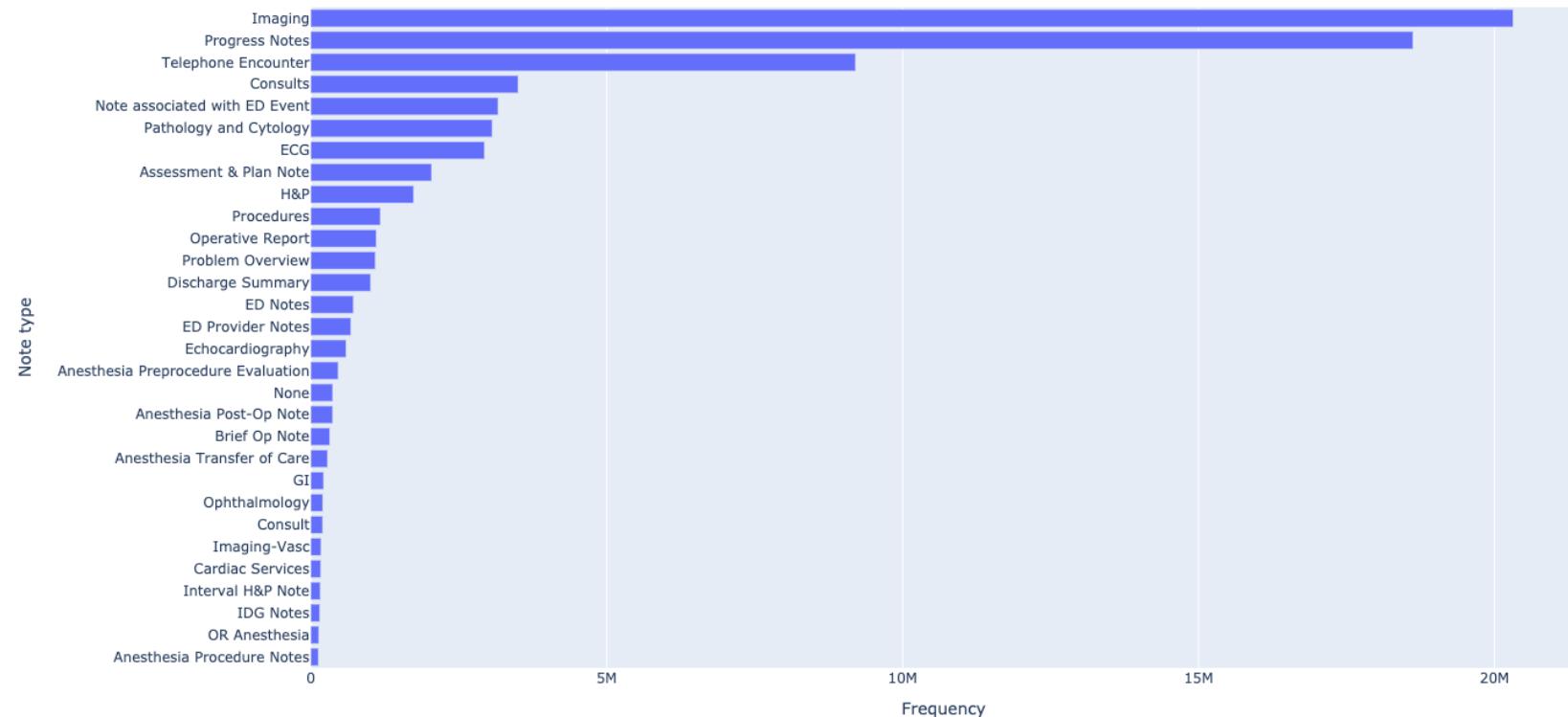
Deidentified clinical notes: statistics

100+ million clinical reports available, mostly from 2012.

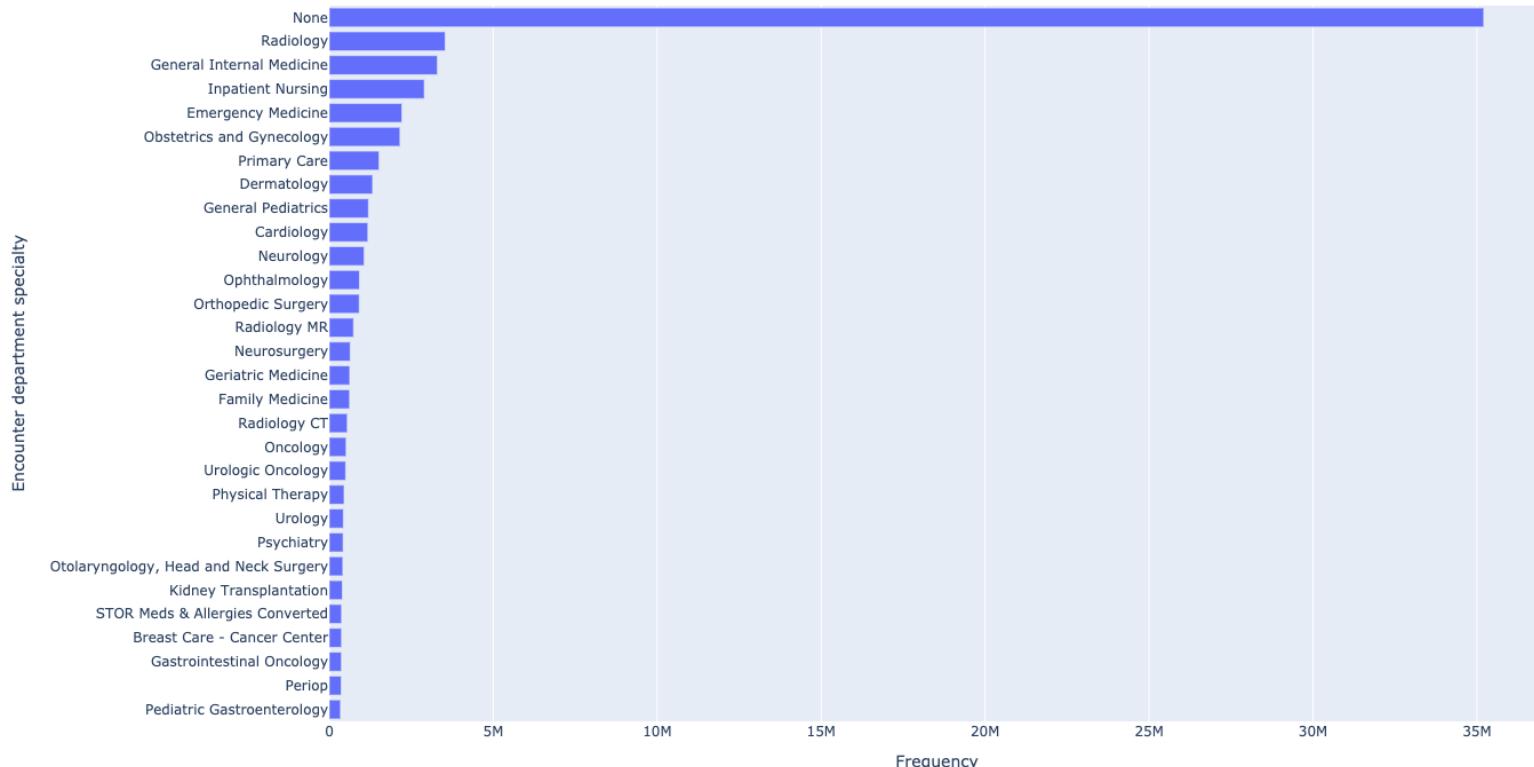
One of the largest sources of clinical information
compared to publicly available data sources

Corpus	Size in words
UCSF redacted notes	39.1B
PMC full-text articles	13.5B
PubMed abstracts	4.5B
MIMIC-III notes	0.5B
English medical Wikipedia	40M
Scientific Papers (SciBERT)	3.17B
English Wikipedia	2.5B
BooksCorpus	0.8B

Clinical reports distribution at UCSF



Clinical reports distribution at UCSF



Clinical notes: sample

HISTORY OF PRESENT ILLNESS Identification: ***** ***** is a 78 y.o. woman with a left neck mass cytologically positive for **B cell lymphoma** here to discuss open biopsy.

...

Medications Current Medications **ACETAMINOPHEN (TYLENOL EXTRA STRENGTH ORAL)** Take 500 mg by mouth continuous prn. ...

ALLERGIES No Known Allergies

Social History Marital status: Divorced ... Smoking status: **Former Smoker** Packs/day: 0.25 Years: 20.00 Types: Cigarettes Start date: 06/10/1959 Quit date: 02/14/1986 **Smokeless tobacco:** Never Used Comment: Don't resume smoking. Alcohol use No Drug use: No Social History Narrative Originally from *****. Family History No history of thyroid disease, thyroid cancer, or any head and neck cancer in the family. **Family History** Diabetes Mother Diabetes **Sister Prostate cancer** Father Died at 83

...

ASSESSMENT Left level 1 b 2.7 cm lymph node: Target for open biopsy. Significant Past Medical History: atrial fibrillation Blood thinners: warfarin Current smoking: none PLAN Anesthesia appointment with EKG. Cardiology clearance for general anesthesia/surgery. **Hold warfarin prior to surgery.** Schedule the patient for: **Left neck open lymph node biopsy.** I explained to the patient the risks, benefits, and possible alternatives of this procedure. The risks include but are not limited to: Bleeding, hematoma, seroma, infection, abscess, unwanted incisional scar, numbness of the face, and neck from sensory cutaneous nerve injury, and temporary or permanent cranial nerve injury to the marginal mandibular, lingual, or hypoglossal nerves. In addition, the risks include complications arising from general anesthesia and hospitalization such as stroke, heart attack, pneumonia, pulmonary emboli, DVT, UTI, Clostridium difficile, and others. The patient expressed understanding and is interested in pursuing this treatment at this time.

Clinical notes: sample

Impression: Metastatic ER negative, HER2 negative breast cancer on nab paclitaxel and pembrolizumab for ***** ***** presents today for follow up, she is doing well and but has clinical or radiological evidence of progression for which she is to be treated with chemo-immunotherapy for her MSI high metastatic disease. The imaging studies were reviewed today and compared to prior scans. Labs are in range for continuation of treatment

Plan: Continue with therapy until progression or intolerable side effects Imaging guided by symptoms or every 3-4 months, longer intervals if stable.

SUBJECTIVE ***** is a 44 y.o. female with Lynch syndrome with early stage colon cancer and endometrial cancer and newly diagnosed metastatic breast cancer to liver and nodes. She started pembrolizumab and abraxane on 03/11/19 and presents today for cycle 1 day 8. At today's visit she reports worsening GERD. Prilosec helps. Taking prn. Also reports nausea for which ativan works well. She reports fatigue for 3 days after treatment. She reports right axillary pain to shoulder blade on the right, much better since last week! Also right lumbosacral pain is stable. She reports new aching in her arms and legs. Continues flexeril and MS contin as well as oxycodone prn. She reports numbness/tingling in toes and fingertips that is transient. She also reports leg cramping at night. Off xarelto still since recent biopsy- was on it for non-occlusive jugular clot.

...

ASSESSMENT & PLAN ***** is a 44 y.o. female with Lynch syndrome with early stage colon cancer and endometrial cancer and newly diagnosed metastatic breast cancer to liver and nodes. She started pembrolizumab and abraxane on 03/11/19 and presents today for cycle 1 day 8.

Metastatic breast cancer - axillary pain improved which is hopeful for early treatment response - Continue abraxane every week (continuous) and Pembrolizumab every 3 weeks. Monitor TSH - Due to poor peripheral access she will need a port, hopefully she can get it next week Jugular clot - restart xarelto, stop for port placement per IR Colon cancer Endometrial cancer Nausea/GERD due to chemo - Continue ativan and compazine prn. Allergic to zofran - Rx prilosec 40mg qd Peripheral neuropathy due to Taxane, new and mild - Monitor Myalgias due to Taxane - Rx mag oxide 400mg qhs Pain - Continue work with pain mgmt - Continue MS Contin 15mg BID, oxycodone prn, flexeril, and ativan - continue stool softener to prevent opioid induced constipation F/u 2 weeks then every 3 weeks

Sample social work report

This 48 y/o woman with stage IV colon cancer lives with her 24 y/o daughter and 4 y/o granddaughter in *****, *****. She also has a 31 year-old daughter with psychiatric disabilities, and a 25 y/o son who is in the ***** in ***** state, who recently was diagnosed with colon cancer himself, and is in treatment. Her mother lives "2 hours away" from her and is not supportive. The daughter that she lives with is supportive but busy with her 4 year old and work. Finances are difficult in the household: Patient has a small fixed income from SSDI and SSI, and has been working part time in food service at the local state hospital when she feels well enough. Patient has a car that she cannot afford to fix and has to rely on others for transportation. She also states that **she sometimes goes without her medications or missing primary care MD visits because she cannot afford her co-pays**. Patient has HMO insurance coverage through her employer that is primary, with Medicare secondary. She is unclear whether she has Medi-Cal coverage that is active now.

...

She states that she is **not sure that she wants to undergo chemotherapy again, due to poor quality of life during past treatments**, and not yet having clear information about the potential benefits of additional chemotherapy in the future.

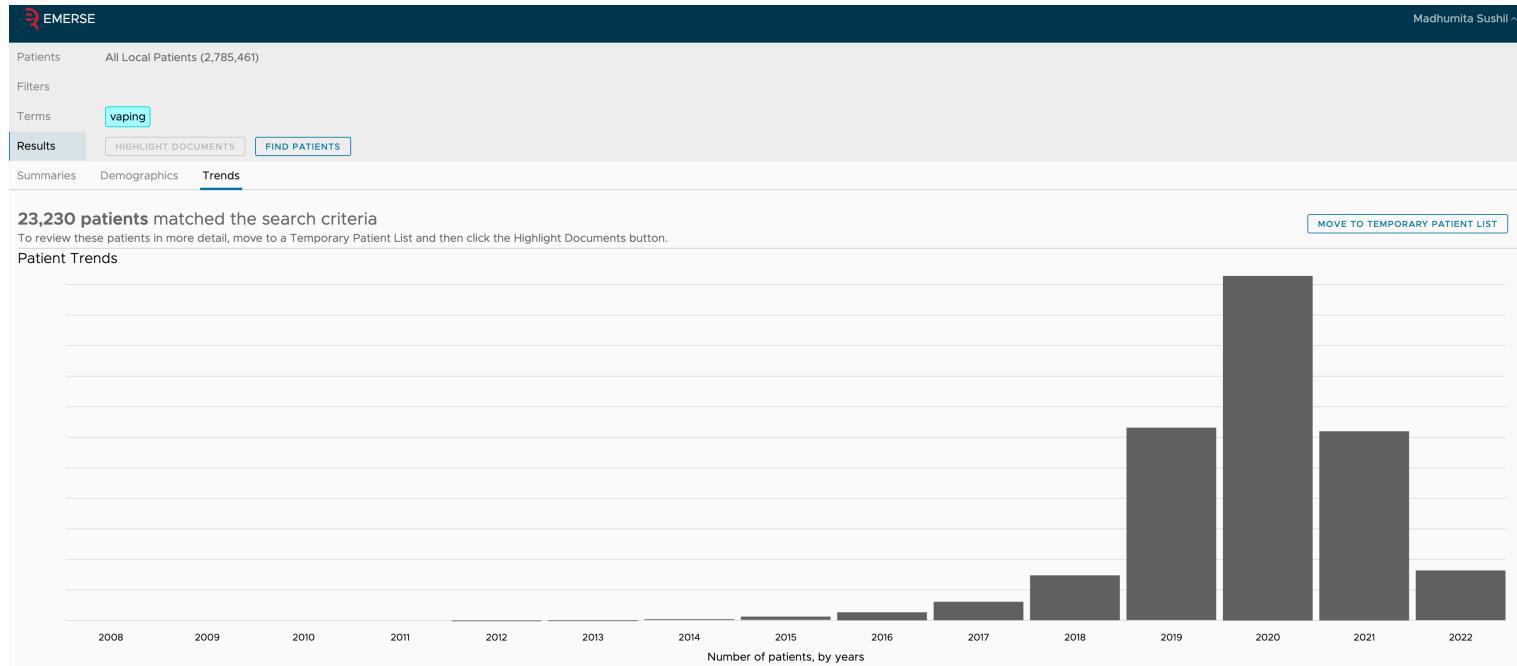
...

Patient has had many psychosocial stressors throughout her life, including early loss of a parent, abandonment by a parent, and several male partners who have been controlling, emotionally abusive and at times physically abusive. There is no abusive partner currently. Her partner of 17 years recently ended their relationship in a way that she describes as very painful. Patient reports missing the companionship and feeling isolated. She would like to return to work for that reason, to see friends. Patient reports a **history of depression and anxiety and has recently been taking Zoloft**, but does not have a history of counseling. She reports **one suicide attempt "six or seven years ago"** that she now sees as more of a suicidal gesture, and denies any other subsequent suicidal ideation. She states that she has been "fighting for her life" and finds it ironic that she would have tried to take her life. She did have one visit with a visiting psychiatrist at work (not on regular staff), whom she reports was not helpful, and not available for ongoing talk therapy or medication management. **She is interested in finding counseling if it is available close to home and is affordable.**

...

SW advised patient to ***** ***** ***** to check her Medi-Cal eligibility so that she can have her co-pays billed to Medi-Cal, to ensure that she can purchase needed medications and attend primary MD visits.

Exploring clinical reports based on keywords



Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform. 2015 Jun;55:290-300. PMID: 25979153.

Common paradigms for clinical information extraction with Natural Language Processing

Rule-based approaches

Pre-defined set of rules (e.g., a template) for processing information from text based on domain knowledge. Examples:

If *vaping* is present in text and *no* is not mentioned before vaping in the text: assume that a patient vapes

If *ER is positive* is mentioned in pathology report, assume that the patient is ER+

If a clinical report mentions *an infection-related term* (from a predefined list), and *at least 2 terms related to inflammatory response* (also from a predefined list), then assume that the patient has *sepsis*.

Limitations

- Requires extensive domain knowledge for the construction of precise, clinically-relevant rules
- Precision and recall tradeoff is quite critical
 - Low precision when trying to generalize more
 - Low recall when the rules are highly specific for prioritizing correct matches
 - Balance becomes tricky when language variability is high
 - Often loses precision when longer contexts are required for inference
 - For example: *ER is positive* could be mentioned either as a test result or within a template of text on how to interpret the results.
- Often not generalizable to different medical centers, different specialties, and different tasks.

Supervised (traditional) machine learning

- Manually labeled datasets are created to provide examples of tasks and labels
- An AI algorithm automatically infers relevant patterns for a specific task from these examples
- Directly applies the automatically learned patterns to unseen data
- More variable the dataset, the more the generalization
- Example:

ER+ : positive
ER and PR +: positive
ER negative: negative
ER positive: positive
ER positive if ... : unknown
ER- : negative

Manually labeled data curation

ER and (+ or positive) and not (if) -> positive
(if) -> unknown
(ER) and (-) or (negative) -> negative

Example of internally learned representation

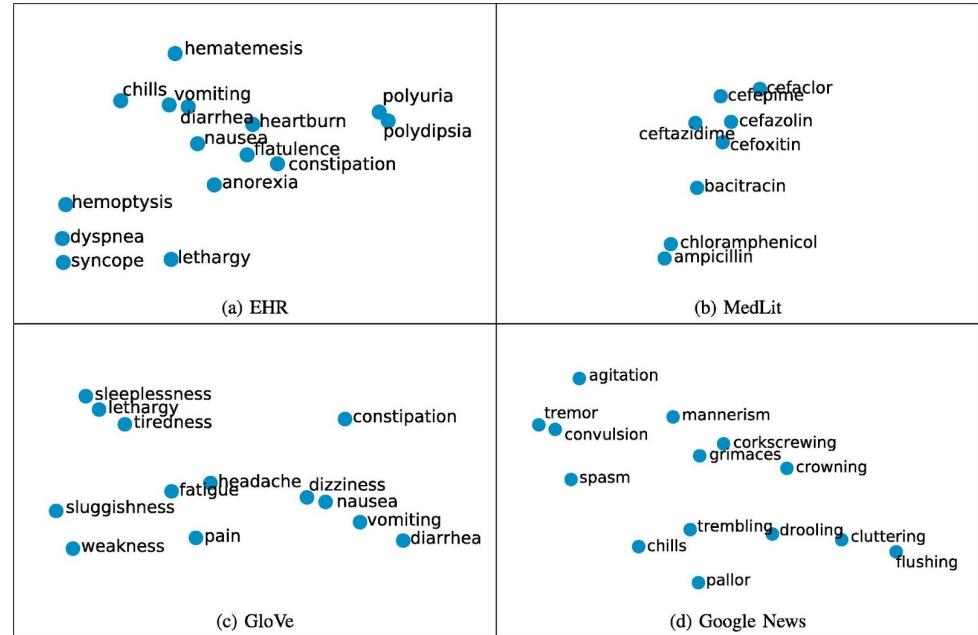
... ER/PR+ ...
Unseen text → positive

Limitations

- Manually labeling data is both time-consuming and expensive
- Large manually labeled datasets are required for generalization, small datasets can cause overfitting
- Can require extensive hyperparameter tuning for developing better models
- Information learned by the models is not entirely interpretable

Supervised deep learning

- A special case of supervised machine learning: using neural networks
- Often learns distributed, low-dimensional data representations to better encode semantics.
 - Meaning of a term is assumed by the context it is used in: terms used in similar contexts get similar representations
- Supports transfer learning from one task to another



A comparison of word embeddings for the biomedical natural language processing,
Yanshan Wang, Sijia Liu, Naveed Afzal, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Paul Kingsbury and Hongfang Liu
Journal of Biomedical Informatics, Volume 87, 2018

Limitations

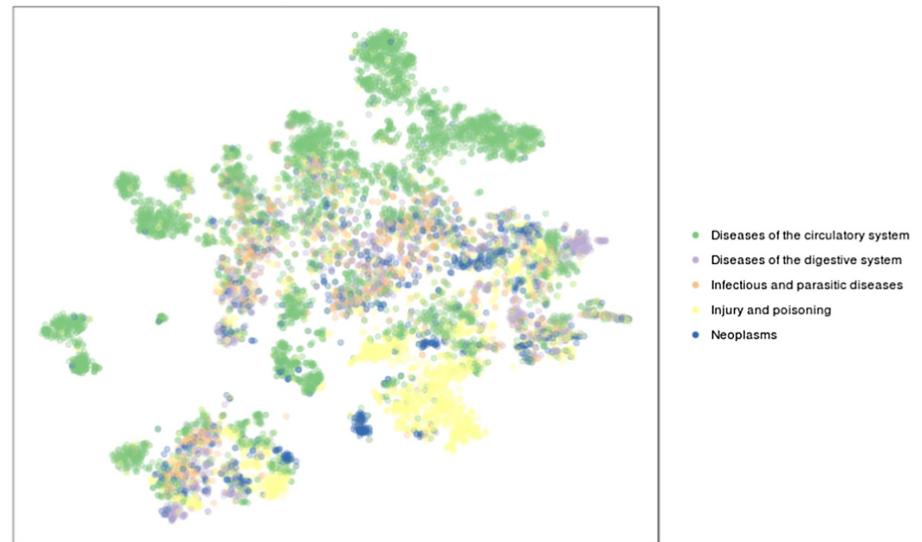
- Usually more prone to overfitting than traditional machine learning algorithms
 - Hence more data hungry
 - Hence more expensive to create large datasets for deep learning algorithms.
- Usually more computationally expensive to train a good model
- Usually less interpretable

Unsupervised learning

- Allows automatic discovery of coherent groups in the data without prior labeling
- Distance-based methods are often used for finding these clusters
- Supports exploratory analysis

Examples:

- Topic modeling: automatically identifying topics of discussion from large textual corpora
- Finding similar patients: clustering patients based on their symptoms, diagnoses, medications, etc.



Sushil, Madhumita, et al. "Patient representation learning and interpretable evaluation using clinical notes." *Journal of biomedical informatics* 84 (2018): 103-113.

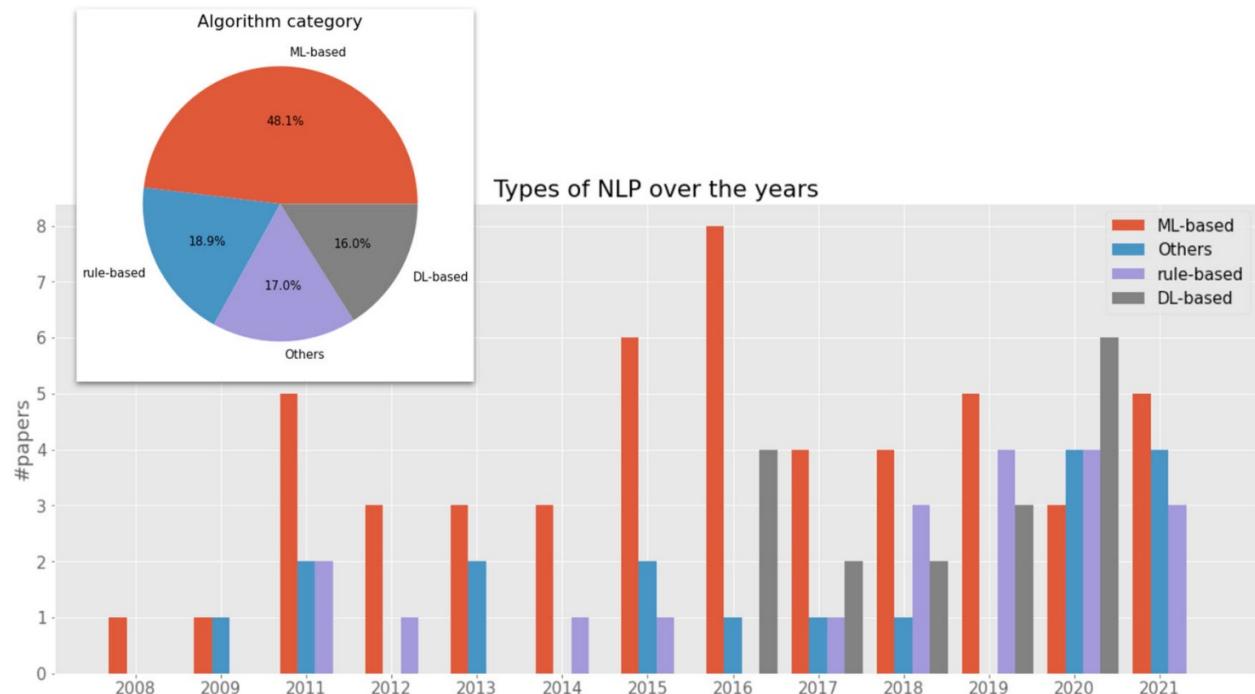
Limitations

- The number of groups often needs to be pre-determined
- Frequently difficult to interpret the obtained groups
- No guarantee that the resulting groups would follow human expectations

NLP Methods overview in real-world studies

Machine learning-based methods are the most popular

Deep learning-based methods have become more popular in the recent years



Transfer learning via pretrained language models

Aims to learn generalized semantic representations of text from large unlabeled corpora and thereby transfer the automatically learned linguistic knowledge to NLP tasks.

The learning objective is often that of (some form of) language modeling, i.e., to predict a missing word in text from the nearby context

Removes the need to re-learn the syntax and semantics of words used in text

The model can instead focus on learning the NLP task instead of the language

Hence, makes downstream tasks more sample efficient

MA

Input: He is on a statin now and we think [...].
Prompt: Create a list of medications. -"statin"
Input: The patient takes coumadin 5 mg [...].
Prompt: Create a list of medications.



- Statin
- Coumadin (5 mg)

<https://chat.openai.com/>

Promises and Limitations

- Promising for zero-shot and few-shot learning
- However, extremely resource intensive, especially for real-world deployment
- Stochastic, and failure is difficult to predict
- Complex to update the existing models for better clinical representation
- Uninterpretable

Case-studies involving natural language processing for clinical research

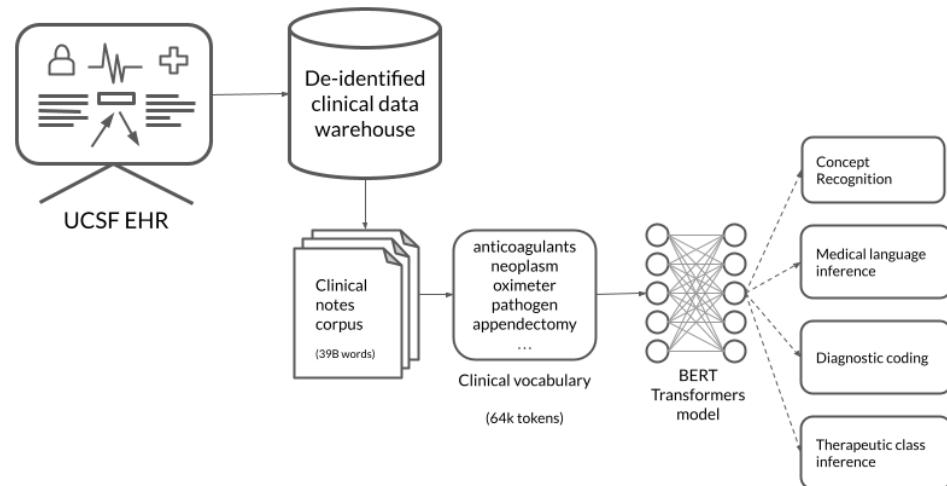
Clinical language modeling for transfer learning

Developing clinical language model from UCSF reports

Custom language model trained from scratch on
75 million de-identified clinical notes at UCSF

Houses a UCSF-specific clinical vocabulary

Trained to learn the general structure and
semantics of clinical rhetoric

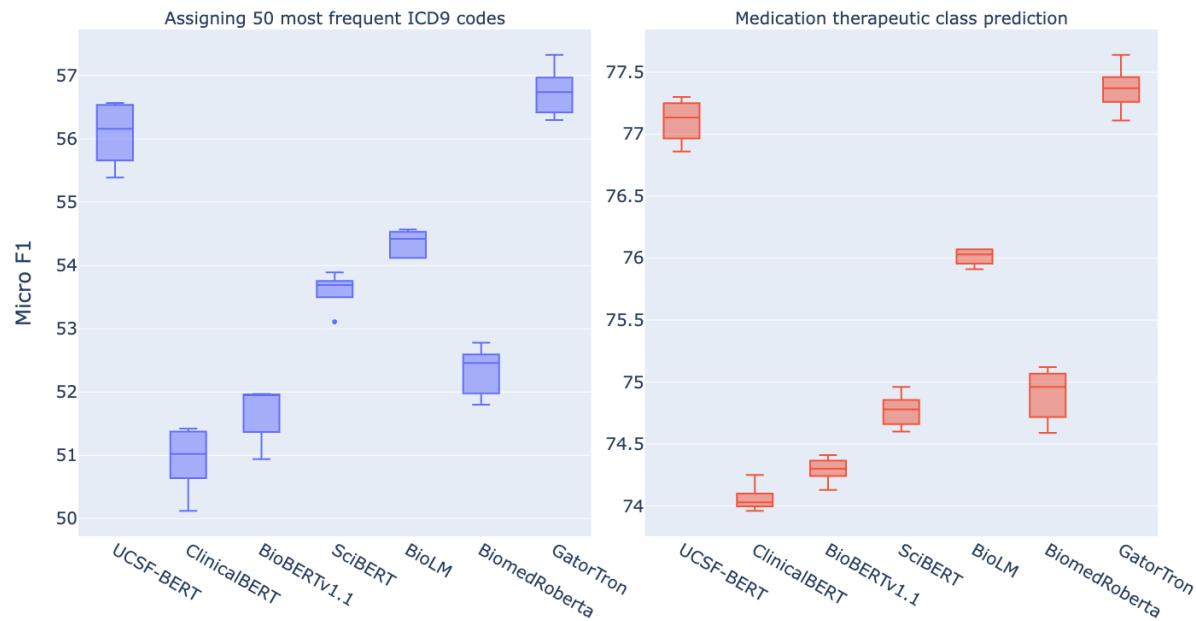


Sushil, Madhumita, et al. "Developing a general-purpose clinical language inference model from a large corpus of clinical notes." *arXiv preprint arXiv:2210.06566* (2022).

Developing clinical language model from UCSF reports

Excellent performance on UCSF-specific NLP tasks, while being much smaller in size than publicly available models.

Transferable to most NLP tasks at UCSF



Sushil, Madhumita, et al. "Developing a general-purpose clinical language inference model from a large corpus of clinical notes." *arXiv preprint arXiv:2210.06566* (2022).

Serious adverse event detection with Natural Language Processing

Clinical reports as real-world evidence for treatments

- Clinical trial efficacy provides an incomplete story due to disparities between trial participants and real-world population
- EHR data consists of wide-scale observational data on the real-world use of drugs and devices
- Clinical notes provide fine-grained information from rhetoric describing the timeline of events and physician interpretations
- Use of clinical reports allows us to monitor: a) current and past treatments, b) events of treatment discontinuation, and c) the reason for discontinuation

Data annotation for adverse event relation extraction

- *History of Present Illness* of 927 patient reports for IBD were manually annotated:
 - Adverse event mentions, hospitalization mentions, and biologic mentions identified with automated rule-based methods
 - Manually added relations:
 - **Biologic before hospitalization:** The last biologic that a patient was on before hospitalization
 - **Reason for hospitalization:** Whether the patient was hospitalized due to the given adverse event

***** is ***** 28 y.o. male with a history of ileocolonic Crohn's disease diagnosed in June 2007 here for follow up. He was last seen September 2016. He was diagnosed with Crohns in 2007 and he had been on azathioprine since August 2010, but continued to have ongoing symptoms. MRI in January 2011 showed evidence of marked thickening^{**} involving the terminal ileum and cecum with surrounding stranding and hyperemia^{**}. The extent and distribution had been unchanged^{**} when compared to previous imaging from 2009 and 2010. In January 2011, the decision was made to switch the patient over to Remicade^{med:Infliximab} and MTX in March 2011. At his visit in March 2011, after receiving the induction dose of Remicade^{med:Infliximab}, the patient had reported significant improvement in his symptoms. He self discontinued MTX in October 2012. Re staging MRE showed improvement in the inflammation^{**}. At his visit in March 2014 he reported feeling well. He denied bowel symptoms and was tolerating Remicade^{med:Infliximab} infusions without difficulty. Denied extra-intestinal manifestations of inflammatory bowel disease including, joint/muscle pain^{**}, rash^{**}, oral ulcerations^{**}, or ocular symptoms. He had recently completed college and plans to work for local firm as civil engineer. Reported increase in alcohol intake, usually <2 drinks a week, in celebration of graduation. His LFTs were checked subsequently and had normalized. At his visit in 07/05/14, he reported feeling the best he had felt for a long time. Having October 01 BMs daily, usually formed, no blood or mucus, nocturnal sx, urgency. Denied fever, chills, weight loss, nausea, vomiting, heartburn. He was mainly eating whatever he wants. Eating healthy, low sugar, little processed foods. Tolerating high fiber foods including salad, fresh fruits and vegetables. He did report a red rash^{**} at #10 right and left cheeks for the last two years that is not itchy or painful. He had not seen dermatology. He was planning to move to **** this **** and will be slowly transitioning to live there full time^{**} to attend a graduate program in theology. He was not sure if he will be here for his next routine visit in 6 months and was planning to establish care with Dr. ***** in ****. Colonoscopy on 07/02/16 showed Crohn^{**}'s disease in endoscopic and symptomatic remission^{**} on Remicade^{med:Infliximab} 5 mg/kg q 8 weeks. Pathology was negative. At his last visit in September 2016 he was having symptoms of nausea^{**}. CT September 2016 at **** showed inflammation^{**} at the IC valve and ileum with dilation^{**} of the distal small bowel loops. He was hospitalized^{med:IBD_actv} at **** where re-read showed partial SBO^{med:IBD_actv} with transition point at the IC valve. He was tried with IV steroids with improvement. At his visit we planned a taper for his prednisone, checking a trough of Remicade^{med:Infliximab} and referral to colorectal surgery. Interval history He saw Dr ***** in October 2016 and she felt that there was an active component to the stricture^{**} and recommended medical therapy. He remains on Remicade^{med:Infliximab} every 8 weeks and has been feeling well. Had food poisoning^{**} in December 2017 which was one day, shortlived. Has recovered completely. He is here today to check in, no issues. Getting married^{**} soon and brings fiance with him.

Automated serious adverse event relation classification

Hierarchical BERT modeling while under-sampling negatives for tasks:

1. Relation classification between specific entity groups:
 - Medication before hospitalization
 - Hospitalization for serious adverse event
 - Medication before hospitalization for adverse event
2. Binary document classification:
 - Is serious AE mentioned in the report?

Task	Model	Macro F1	TP	TN	FP	FN
Medication before hospitalization relations	H-UCSF-BERT	0.62	63	1989	173	105
	XGBoost	0.51	49	1672	490	119
	BiLSTM	0.48	51	1670	492	117
Hospitalization for SAE relations	H-UCSF-BERT	0.62	79	9603	421	16
	H-UCSF-BERT + only nearby SAEs	0.68	34	1078	41	61
	BiLSTM + only nearby SAEs	0.48	7	1091	28	8
Medication before hospitalization for adverse event relation (triples)	H-UCSF-BERT + only nearby SAEs	0.61	141	7790	619	178
	BiLSTM + only nearby AEs	0.50	11	7953	456	308
	AutoML: NeuralNetFastAI_BAG_L1 + only nearby AEs	0.58	25	7362	1047	294
Is serious AE in report?	H-UCSF-BERT	0.66	26	36	18	13
	CatBoost_BAG_L1	0.63	36	31	23	3

Next steps

Improving the generalizability of the approach to other types of reports, diseases, and medical centers

Expanding upon the definition of *serious*

Automated resolution of the identified adverse event terms to MedDRA codes

Exploring previously unknown adverse event for hypothesis generation

Moving to the inference of causal relations

Uncovering breast cancer phenotypes with Natural Language Processing

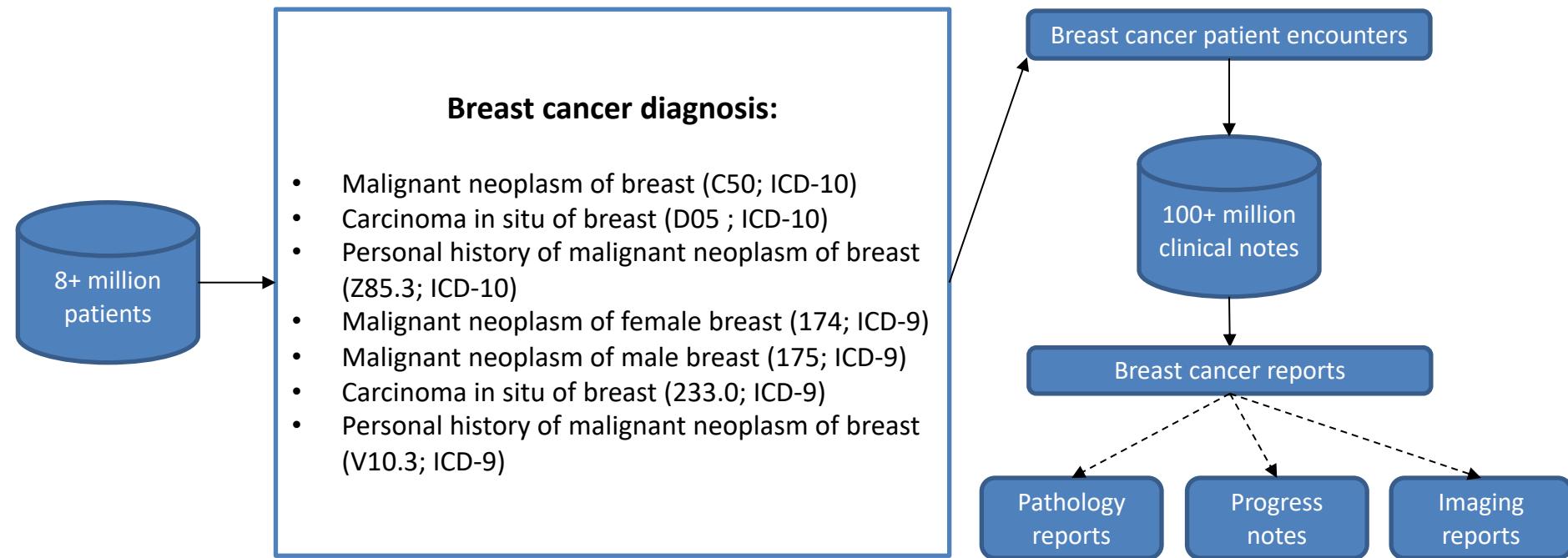
Integrating structured and unstructured data

Motivation: Clinically-relevant information for breast cancer is isolated to textual reports:

- Estrogen Receptors (ER) status
- Progesterone Receptors (PR) status
- HER2 gene amplification
- Tumor grade
- Tumor stage
- Family and social history
- Disease history and progression

Integrating structured information (demographics, diagnosis codes, medications, lab tests, procedures) with fine-grained information from notes provides more complete phenotypes

Retrieving clinical reports for breast cancer patients



Extracting relevant breast cancer pathology: creating labeled pathology dataset

The test for estrogen receptors is positive. There is variable (ranging from weak to strong) nuclear staining in ~70% of tumor cells.

Internal positive control is present.

The test for progesterone receptors is positive. There is moderate to strong nuclear staining in ~80% of tumor cells. Internal positive control is present.

Result of **** test: This carcinoma is negative for **** oncoprotein over-expression.

An immunohistochemical assay was performed by manual morphometry on block **** using the **** monoclonal antibody to **** oncoprotein. The staining intensity of this carcinoma was 1 on a scale of 0-3.

Carcinomas with staining intensity scores of 0 or 1 are considered

FINAL PATHOLOGIC DIAGNOSIS

****. **** lymph node, left axillary, biopsy:

One lymph node with no tumor identified (0/1). See comment.

B. Left breast, partial mastectomy:

1. Invasive ductal carcinoma, **** grade 2, margins negative for tumor.
2. Ductal carcinoma in situ, intermediate grade.
3. Non-proliferative fibrocystic change.

See comment.

C. Sentinel lymph node, left axillary, biopsy:

One lymph node with no tumor identified (0/1).

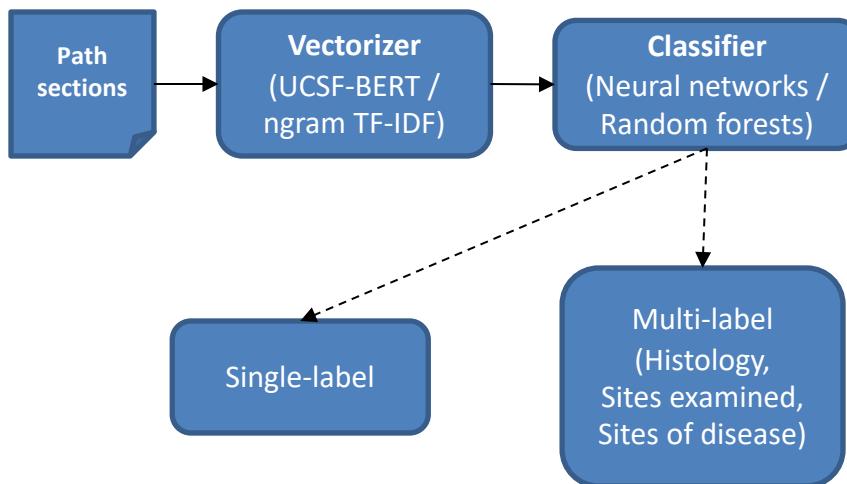
Sites examined	<input checked="" type="checkbox"/> Left Breast ^[g] <input checked="" type="checkbox"/> Left LN ^[z] <input type="checkbox"/> Right Breast ^[x] <input type="checkbox"/> Right LN ^[c] <input type="checkbox"/> Other tissues ^[v] <input type="checkbox"/> Unknown ^[b]
Sites of disease	<input checked="" type="checkbox"/> Left Breast ^[y] <input type="checkbox"/> Left LN ^[i] <input type="checkbox"/> Right Breast ^[o] <input type="checkbox"/> Right LN ^[p] <input type="checkbox"/> Other tissues ^[j] <input type="checkbox"/> None ^[k] <input type="checkbox"/> Unknown ^[l]
Histology	<input type="checkbox"/> No malignancy ^[n] <input type="checkbox"/> LCIS ^[m] <input checked="" type="checkbox"/> DCIS <input checked="" type="checkbox"/> Invasive ductal <input type="checkbox"/> Invasive lobular <input type="checkbox"/> Medullary <input type="checkbox"/> Mucinous <input type="checkbox"/> Tubular <input type="checkbox"/> Papillary <input type="checkbox"/> Metaplastic BC <input type="checkbox"/> Mixed <input type="checkbox"/> Carcinoma NOS <input type="checkbox"/> Unknown
LN involvement	<input checked="" type="checkbox"/> 0 involved <input type="checkbox"/> 1-3 involved <input type="checkbox"/> 4-9 involved <input type="checkbox"/> 10+ involved <input type="checkbox"/> Unknown
Biopsy type	<input type="checkbox"/> Biopsy <input type="checkbox"/> Lumpectomy <input checked="" type="checkbox"/> Mastectomy <input type="checkbox"/> Unknown
ER	<input type="checkbox"/> Low positive <input checked="" type="checkbox"/> Positive <input type="checkbox"/> Negative <input type="checkbox"/> Unknown
PR	<input checked="" type="checkbox"/> Positive <input type="checkbox"/> Negative <input type="checkbox"/> Unknown
HER2	<input type="checkbox"/> Positive <input checked="" type="checkbox"/> Negative <input type="checkbox"/> Equivocal <input type="checkbox"/> Equivocal Positive <input type="checkbox"/> Equivocal Negative <input type="checkbox"/> Unknown
Max grade	<input type="checkbox"/> 0 <input type="checkbox"/> 1 (Low) <input checked="" type="checkbox"/> 2 (Intermediate) <input type="checkbox"/> 3 (High) <input type="checkbox"/> Unknown
LVI	<input checked="" type="checkbox"/> Present <input type="checkbox"/> Absent <input type="checkbox"/> Unknown
Margins	<input type="checkbox"/> Positive margin <input checked="" type="checkbox"/> Less than 2mm <input type="checkbox"/> More than/eq to 2mm <input type="checkbox"/> Unknown
DCIS Margins	<input type="checkbox"/> Positive margin <input checked="" type="checkbox"/> Less than 2mm <input type="checkbox"/> More than/eq to 2mm <input type="checkbox"/> Unknown

Pathology dataset statistics

- Number of annotated path reports:
400 (Round 1) + 250 (Round 2) + 100
(Round 3; 3 independent annotators)
- 100 notes annotated by 2 annotators
for inter-annotator agreement,
thereby 1 annotator per note

Task	Inter-annotator agreement (Krippendorff's alpha)
Biopsy type	0.80
Num lymph nodes involved	0.89
ER	0.85
PR	0.90
HER2	0.80
Grade	0.85
LVI	0.97
Margins	0.93
DCIS Margins	0.77
Histology (Multilabel)	0.82
Sites examined (Multilabel)	0.79
Sites of disease (Multilabel)	0.85

Developing pathology extraction pipelines



Task	Random F. (Macro F1)	LSTM Attn (Macro F1)	UCSF-BERT (Macro F1)
Pathology type	0.93	0.93	0.91
Biopsy type	0.81	0.80	0.88
ER	0.79	0.93	0.82
PR	0.87	0.93	0.76
HER2	0.64	0.97	0.80
Grade	0.42	0.60	0.24
LVI	0.63	0.90	0.68
Margins	0.37	0.60	0.39
Histology (Multilabel)	0.63	0.74	0.43
Sites examined (Multilabel)	0.60	0.71	0.89
Sites of disease (Multilabel)	0.47	0.76	0.72

Extracting entities from progress reports: Creating a comprehensive labeled dataset for few-shot learning



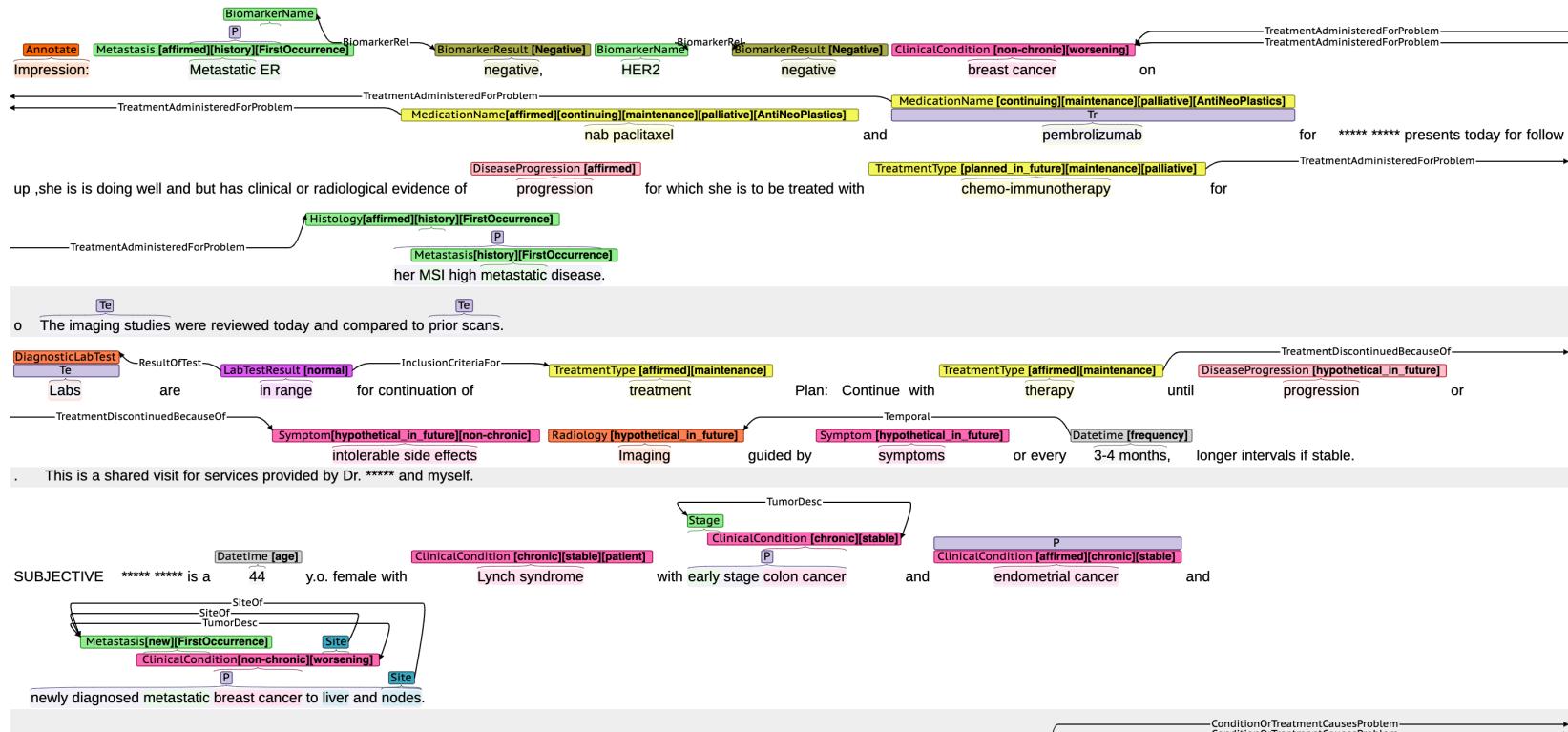
+ Relations

- Temporal relations between a datetime entity and any other entity
- Descriptive relations, for example, relation between a test and its result
- Advanced relations: those requiring implicit or explicit inference
 - TestOrProcedureConductedForProblem
 - TreatmentDiscontinuedBecauseOf
 - TestOrProcedureReveals X
 - ConditionOrTreatmentCausesProblem
 - TreatmentAdministeredForProblem
 - X NotUndergoneBecauseOf Y
 - InclusionCriteria
 - ExclusionCriteria

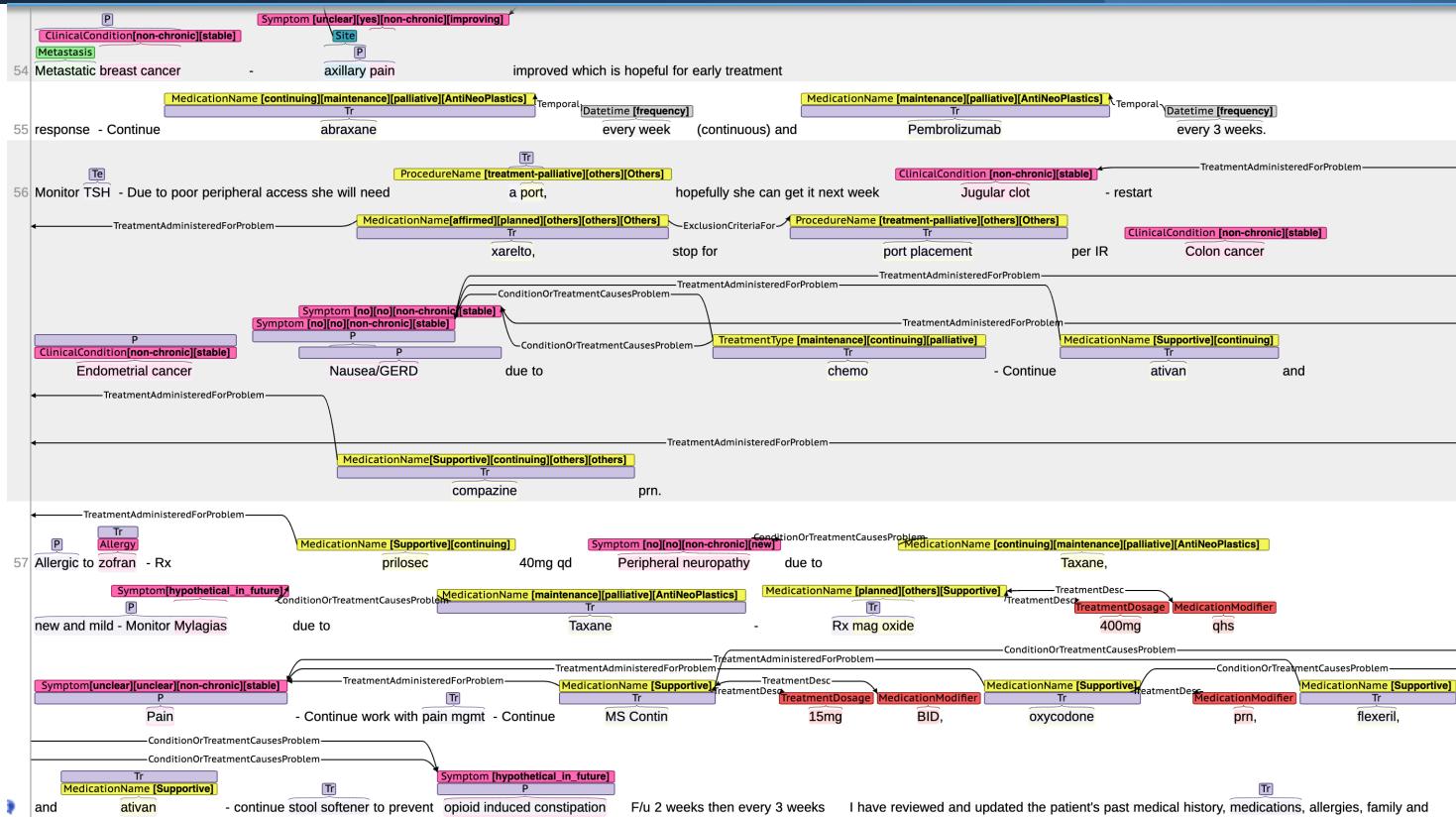
+ Attributes

- Modifiers of the entities. Examples:
 - Negation
 - Experiencer
 - Is a symptom caused due to cancer diagnosis
 - Temporality (history/new; stable/improving/worsening; meds finished/discontinued early?)
 - Cancer episodes
 - Intent of a test or a procedure
 - Type of treatment (adjuvant/neo-adjuvant/maintenance/local, curative/palliative, anti-neoplastic/others)

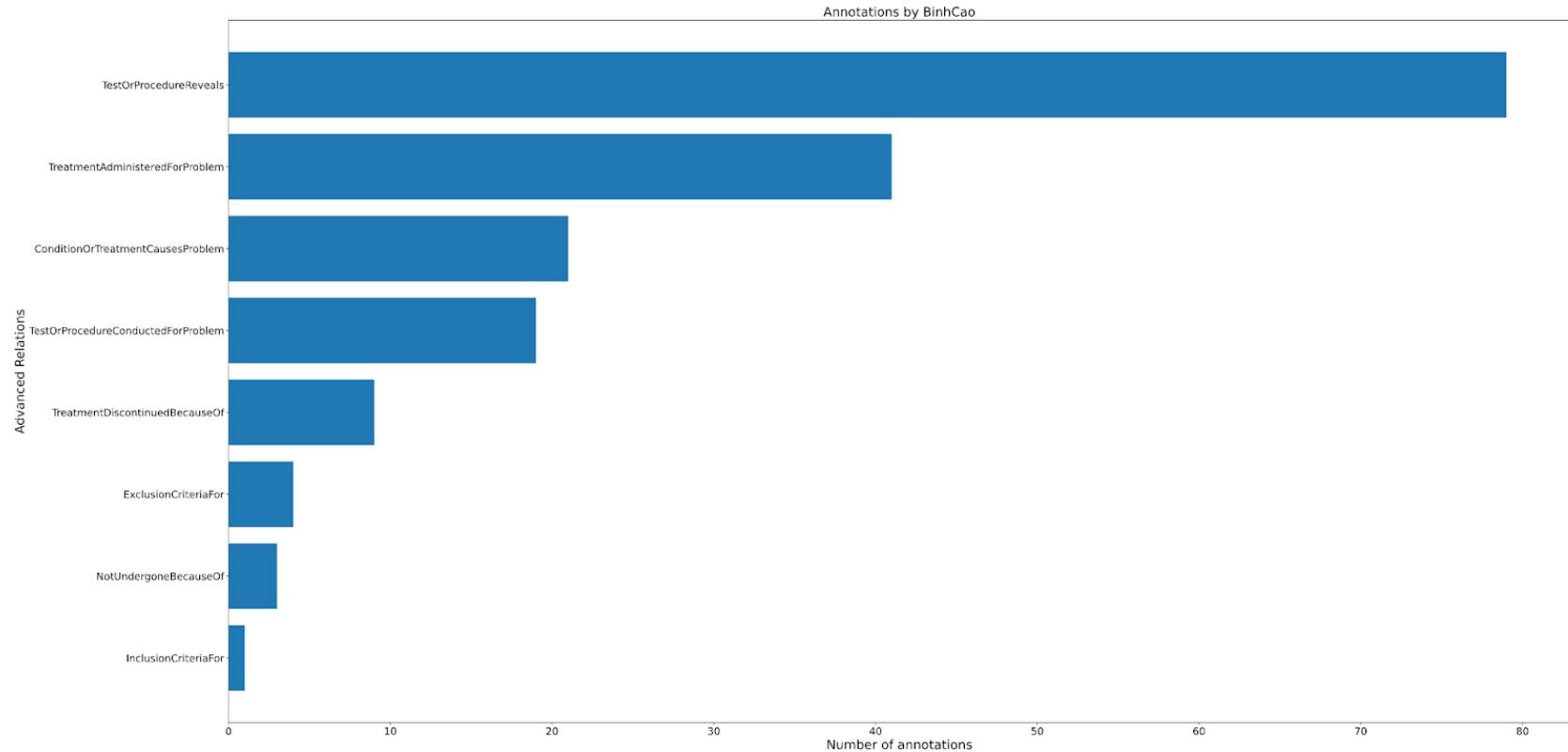
Sample annotation



Sample annotation



Inference-based relation representation across 4 advanced Oncology progress reports



Zero-shot phenotyping with ChatGPT

MA

The patient has not had screening mammograms. She felt a lump in her left breast in February 2017. She did not immediately seek care, but did start an evaluation (through *****) in June 2017. Her work-up and treatment to date are as follows: February 2017 Felt lump in her left breast 06/14/17 Diagnostic mammogram: irregular spiculated mass in left upper outer quadrant corresponding to palpable abnormality; irregular mass in posterior, inferior, subareolar region of right breast Targeted left breast ultrasound: irregular hypoechoic mass (1.8 x 2.3 x 2.7 cm) at 1:30 position, 10 cm from nipple Targeted right breast ultrasound: lobulated hypoechoic mass (1 x 1.1 x 1.2 cm) at 8:00 position, 6 cm from nipple 06/17/17 US-guided left breast core biopsy at 8:00 position, 6 cm from nipple: 0.75 cm grade 3 IDC with ER+ 100%, PR+ 75%, HER2- (IHC 1+; FISH ratio 1.1) US-guided right breast core biopsy at 1:30 position, 10 cm from nipple: 0.6 cm grade 1 IDC with ER+ 100%, PR+ 75%, HER2+ (IHC 2+; FISH ratio 2.0 with HER2 copy # per cell 4.4); intermediate nuclear grade DCIS 07/02/17 Breast MRI: right - irregular homogeneously enhancing mass (2.3 x 1.4 x 1.1 cm) corresponding to right breast index mass (at 8:00 position, 13 cm from nipple); left - irregular homogeneously enhancing mass (2.2 x 3.3 x 2.6 cm) corresponding to left breast index mass (at 2:00 position, 10 cm from nipple); multiple prominent low level 1 left axillary LN with thickened cortex, asymmetric compared to right 07/11/17 US-guided left axillary LN core biopsy (3:00 position, 19 cm from nipple): multiple foci of metastatic adenocarcinoma (largest 1.5 mm) with tumor cells present in adipose tissue adjacent to fibrous tissue, unclear whether this represents extracapsular invasion; HER2- (IHC 1+; FISH ratio 1.2) 08/13/17 Left breast partial mastectomy and axillary lymph node dissection: 6.2 cm grade 3 IDC; ER+ 99%, PR+ >95%, HER2- (1+; FISH ratio 1.0); suspicion for LVI, margins negative, focal high grade DCIS within tumor; 0/8 LN+ (0.7 mm micrometastasis with no extracapsular extension); MP high risk (-0.614) Right breast partial mastectomy and axillary lymph node dissection: 2.0 cm grade 1 IDC with ER+ 99%, PR+ 90%, HER2- (IHC 0; FISH ratio 1.5); 0/8 LN+; MP Low Risk (+0.321) (Surgeon: Dr. *****). Present to clinic with her husband and sister (****). She reports that she has felt a lot healthier since having her breast surgery. She has some mild joint stiffness, in particular in her knees. **** has met with Dr. ***** and Dr. ***** ***** who have both recommended adjuvant chemotherapy (she states that they have specifically recommended TC). She states that when she saw Dr. ***** he recommended that her tumor (presumptively left) be sent for Oncotype. She is not yet aware of a result. She is concerned that the benefits of chemotherapy will not outweigh its long-term risks. Past Medical History: Diagnosis Date Breast cancer Diabetes 08/09/2017 HbA1c of October 20 on 08/09/17 Obesity Past Surgical History: Procedure Laterality Date AXILLARY LYMPH NODE DISSECTION Bilateral 10/09/2017 8 LN ON RIGHT; 12 LN ON LEFT MASTECTOMY, PARTIAL 08/13/2017 WITH BILATERAL REDUCTION Family History Problem Relation Name Age of Onset Breast cancer Maternal Grandmother 58

Give me all cancer biomarkers



The cancer biomarkers for the patient in the case are:



- ER (Estrogen Receptor): ER+ 100% in left breast mass and ER+ 99% in right breast mass
- PR (Progesterone Receptor): PR+ 75% in left breast mass and PR+ 90% in right breast mass
- HER2 (Human Epidermal growth factor Receptor 2): HER2- (IHC 1+) in left breast mass, HER2+ (IHC 2+; FISH ratio 2.0 with HER2 copy # per cell 4.4) in right breast mass, and HER2- (IHC 1+; FISH ratio 1.2) in left axillary lymph node
- MP (Mammaprint): high risk (-0.614) in left breast and low risk (+0.321) in right breast

MA

Can you give me that in a table form?



Sure, here is a table summarizing the biomarkers mentioned in the case:

Biomarker	Left breast biopsy	Right breast biopsy
ER	100% positive	100% positive
PR	75% positive	90% positive
HER2	- (IHC 1+, FISH ratio 1.1)	2+ (IHC 2+, FISH ratio 2.0, HER2 copy # per cell 4.4)

Note: The table is based on the information provided in the case, which may not be comprehensive or fully up-to-date.

More zero-shot phenotyping with ChatGPT

MA

The patient has not had screening mammograms. She felt a lump in her left breast in February 2017. She did not immediately seek care, but did start an evaluation (through *****) in June 2017. Her work-up and treatment to date are as follows: February 2017 Felt lump in her left breast 06/14/17 Diagnostic mammogram: irregular spiculated mass in left upper outer quadrant corresponding to palpable abnormality; irregular mass in posterior, inferior, subareolar region of right breast Targeted left breast ultrasound: irregular hypoechoic mass (1.8 x 2.3 x 2.7 cm) at 1:30 position, 10 cm from nipple Targeted right breast ultrasound: lobulated hypoechoic mass (1 x 1.1 x 1.2 cm) at 8:00 position, 6 cm from nipple 06/17/17 US-guided left breast core biopsy at 8:00 position, 6 cm from nipple: 0.75 cm grade 3 IDC with ER+ 100%, PR+ 75%, HER2- (IHC 1+; FISH ratio 1.1) US-guided right breast core biopsy at 1:30 position, 10 cm from nipple: 0.6 cm grade 1 IDC with ER+ 100%, PR+ 75%, HER2+ (IHC 2+; FISH ratio 2.0 with HER2 copy # per cell 4.4); intermediate nuclear grade DCIS 07/02/17 Breast MRI: right - irregular homogeneously enhancing mass (2.3 x 1.4 x 1.1 cm) corresponding to right breast index mass (at 8:00 position, 13 cm from nipple); left - irregular homogeneously enhancing mass (2.2 x 3.3 x 2.6 cm) corresponding to left breast index mass (at 2:00 position, 10 cm from nipple); multiple prominent low level 1 left axillary LN with thickened cortex, asymmetric compared to right 07/11/17 US-guided left axillary LN core biopsy (3:00 position, 19 cm from nipple): multiple foci of metastatic adenocarcinoma (largest 1.5 mm) with tumor cells present in adipose tissue adjacent to fibrous tissue, unclear whether this represents extracapsular invasion; HER2- (IHC 1+; FISH ratio 1.2) 08/13/17 Left breast partial mastectomy and axillary lymph node dissection: 6.2 cm grade 3 IDC; ER+ 99%, PR+ >95%, HER2- (1+; FISH ratio 1.0); suspicion for LVI, margins negative, focal high grade DCIS within tumor; 0/8 LN+ (0.7 mm micrometastasis with no extracapsular extension); MP high risk (-0.614) Right breast partial mastectomy and axillary lymph node dissection: 2.0 cm grade 1 IDC with ER+ 99%, PR+ 90%, HER2- (IHC 0; FISH ratio 1.5); 0/8 LN+; MP Low Risk (+0.321) (Surgeon: Dr. *****) Present to clinic with her husband and sister (****). She reports that she has felt a lot healthier since having her breast surgery. She has some mild joint stiffness, in particular in her knees. **** has met with Dr. ***** and Dr. ****. **** who have both recommended adjuvant chemotherapy (she states that they have specifically recommended TC). She states that when she saw Dr. **** he recommended that her tumor (presumptively left) be sent for Oncotype. She is not yet aware of a result. She is concerned that the benefits of chemotherapy will not outweigh its long-term risks. Past Medical History: Diagnosis Date Breast cancer Diabetes 08/09/2017 HbA1c of October 20 on 08/09/17 Obesity Past Surgical History: Procedure Laterality Date AXILLARY LYMPH NODE DISSECTION Bilateral 10/09/2017 8 LN ON RIGHT; 12 LN ON LEFT MASTECTOMY, PARTIAL 08/13/2017 WITH BILATERAL REDUCTION Family History Problem Relation Name Age of Onset Breast cancer Maternal Grandmother 58

Give me all cancer biomarkers

MA

Tabulate the history of radiology tests and results for me.



Here is a table of the radiology tests and results as described in the patient's history:

Test	Date	Result
Diagnostic mammogram	06/14/2017	Irregular spiculated mass in left upper outer quadrant corresponding to palpable abnormality; irregular mass in posterior, inferior, subareolar region of right breast
Targeted left breast ultrasound	06/14/2017	Irregular hypoechoic mass (1.8 x 2.3 x 2.7 cm) at 1:30 position, 10 cm from nipple
Targeted right breast ultrasound	06/14/2017	Lobulated hypoechoic mass (1 x 1.1 x 1.2 cm) at 8:00 position, 6 cm from nipple
US-guided left breast core biopsy	06/17/2017	0.75 cm grade 3 IDC with ER+ 100%, PR+ 75%, HER2- (IHC 1+; FISH ratio 1.1)
US-guided right breast core biopsy	06/17/2017	0.6 cm grade 1 IDC with ER+ 100%, PR+ 75%, HER2+ (IHC 2+; FISH ratio 2.0 with HER2 copy # per cell 4.4); intermediate nuclear grade DCIS
Breast MRI	07/02/2017	Right - irregular homogeneously enhancing mass (2.3 x 1.4 x 1.1 cm) corresponding to right breast index mass (at 8:00 position, 13 cm from nipple); left - irregular homogeneously enhancing mass (2.2 x 3.3 x 2.6 cm) corresponding to left breast index mass (at 2:00 position, 10 cm from nipple); multiple prominent low level 1 left axillary LN with thickened cortex, asymmetric compared to right
US-guided left axillary LN core biopsy	07/11/2017	Multiple foci of metastatic adenocarcinoma (largest 1.5 mm) with tumor cells present in adipose tissue adjacent to fibrous tissue, unclear whether this represents extracapsular invasion; HER2- (IHC 1+; FISH ratio 1.2)

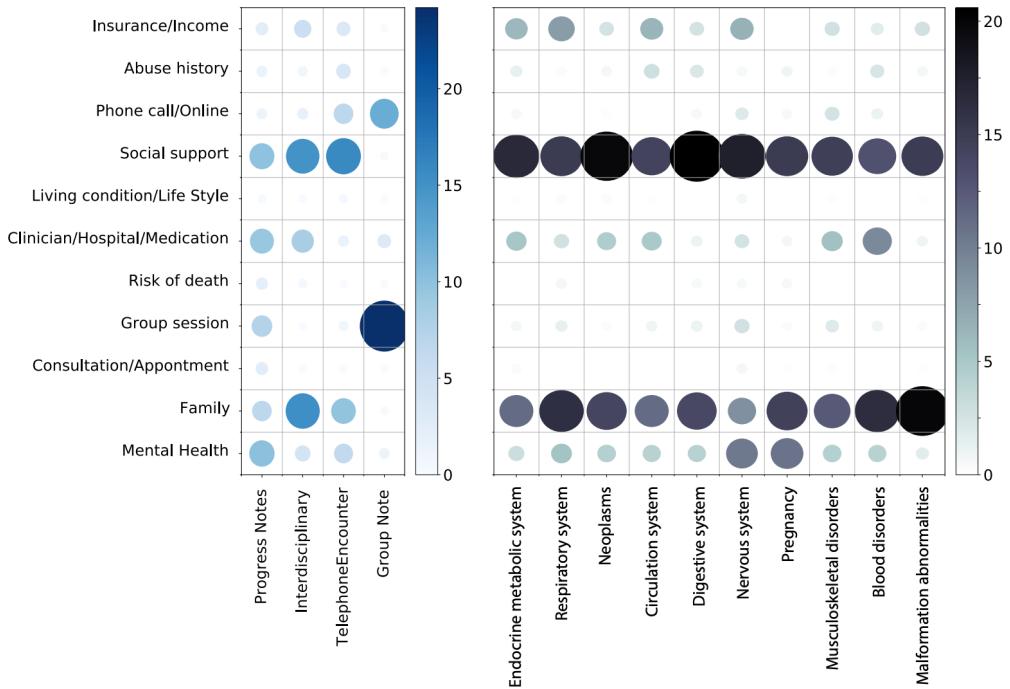
Next steps: exploring advanced inference in Oncology with ChatGPT

- **Symptoms**
 - Symptoms present at the time of cancer diagnosis
 - Symptoms most likely to be caused by the diagnosed cancer
- **Diagnostic tests**
 - Medical indications for each administered test, and the results most relevant for diagnosing cancer
- **Medications/Treatments**
 - Ongoing, new, discontinued treatments
 - Reasons for the administration of a specific drug
 - Reasons for discontinuing a specific drug
 - Potential adverse events for a drug
 - Experienced adverse events due to a drug

Social determinants of health in clinical reports

Topic modeling

- Unsupervised clustering of social work reports at UCSF to identify topics of discussion by note type and disease category
- Clusters are assigned a topic label with manual review
- Frequent administrative discussions, but topics about insurance, abuse, family and mental health are also abundant
- Mental health is a prominent topic in pregnancy and neurological diseases



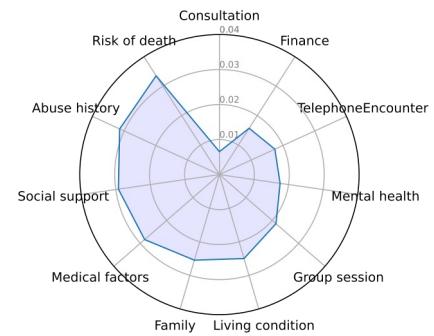
Sun, Shenghuan, Travis Zack, Madhumita Sushil, and Atul J. Butte. "Topic Modeling on Clinical Social Work Notes for Exploring Social Determinants of Health Factors." *arXiv preprint arXiv:2212.01462* (2022).

Association between social work discussion and targeted therapy administration

Using only social work reports, can we predict whether a breast cancer patient received targeted therapy?

- Yes, much better than random, and social factors play an important role in the prediction!
- Adding more social work reports improves the prediction further

Model	AUC	Macro F1
UCSF BERT (8 reports)	0.718	0.623
UCSF BERT (longest report)	0.675	0.599
ClinicalBERT (longest report)	0.627	0.578
BioLM (longest report)	0.671	0.583
Random (stratified)	0.504	0.525



Feature topic importance quantified with ablation analysis

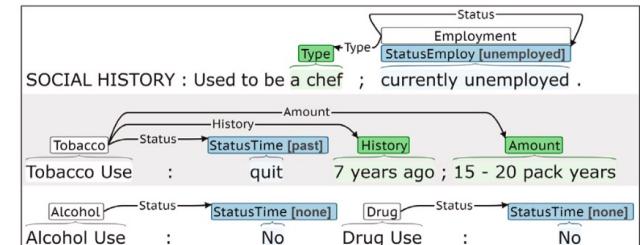
Automatic extraction of SDoH factors from social history

Task A: To extract the mentions of **Substance use** (Alcohol, Drugs, Tobacco), **Employment status**, and **Living status** along with their annotated modifiers from the MIMIC-III subset

Task B: To analyze system generalizability on data from an external institution (University of Washington)

Task C: To transfer the existing models to the external institution with additional annotated data

Event type, e	Argument type, a	Argument subtypes, y_i	Span examples
Substance use (Alcohol, Drug, & Tobacco)	Status ^a	{none, current, past}	"denies," "smokes"
	Duration	-	"for the past 8 years"
	History	-	"seven years ago"
	Type	-	"beer," "cocaine"
	Amount	-	"2 packs," "3 drinks"
Employment	Frequency	-	"daily," "monthly"
	Status ^a	{employed, unemployed, retired, on disability, student, homemaker}	"works," "unemployed"
	Duration	-	"for five years"
	History	-	"15 years ago"
Living status	Type ^a	-	"nurse," "office work"
	Status ^a	{current, past, future}	"lives," "lived"
	Type ^a	{alone, with family, with others, homeless}	"with husband," "alone"
	Duration	-	"for the past 6 months"
	History	-	"until a month ago"



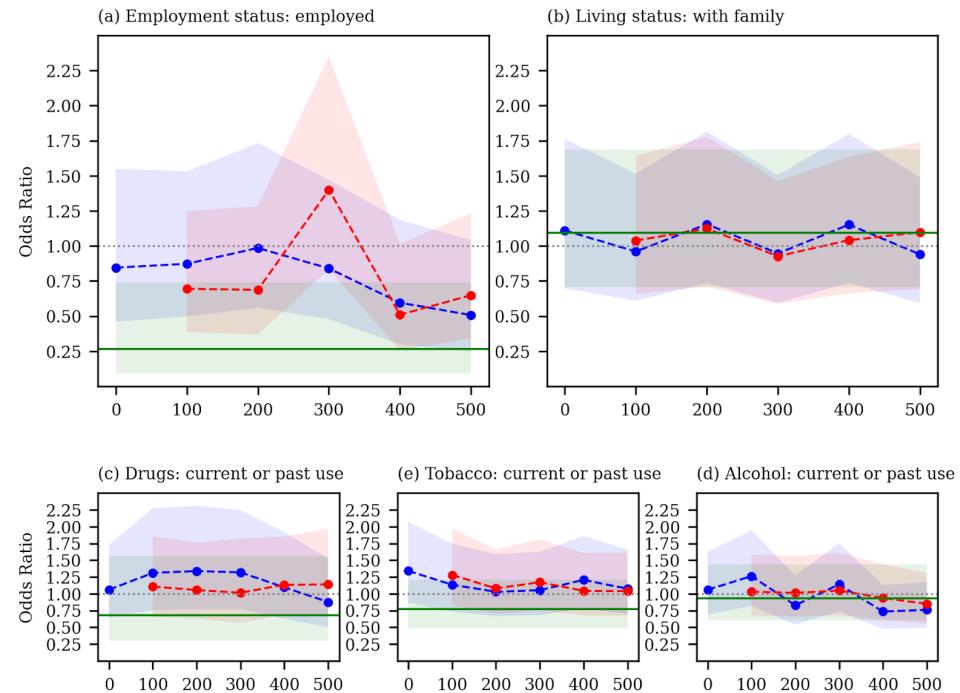
SDoH extraction results

- UCSF BERT model performed the best across all the settings, further improvements obtained later by removing the LSTM layer all together.
- The models suffer significantly (> 15%) in cross-institution generalizability.
- Cross-institution performance improves significantly by adding additional actively annotated data from that institution

	Setting	Training	Testing	Model	P	R	F1
1	Phase A (same domain)	MIMIC _{train+dev} (1504 docs.)	MIMIC _{test} (373 docs.)	*LSTM	0.84	0.65	<u>0.74</u>
2			LSTM + ST _{MIMIC}		0.87	0.64	0.73
3 [†]			*LSTM + fT _{MIMIC@UCSF}		0.83	0.69	0.76
4 [†]			*LSTM + fT _{MIMIC@UCSF} + ST _{MIMIC}		0.85	0.71	0.77
5			BERT _{UCSF} + LSTM		0.81	0.76	0.78
6			BERT _{UCSF} + LSTM + ST _{MIMIC}		0.78	0.75	0.76
7			Shared task mean/median (14 teams)		0.69	0.56	0.60/0.69
8	Phase B <u>(cross domain)</u>	MIMIC _{train+dev} (1504 docs.)	UW _{train+dev} (2010 docs.)	LSTM	0.72	0.42	0.53
9 [†]			*LSTM + fT _{MIMIC@UCSF}		0.72	0.49	0.58
10 [†]			*LSTM + fT _{MIMIC@UCSF} + ST _{UCSF}		0.76	0.46	0.57
11			*BERT _{UCSF} + LSTM		0.68	0.54	<u>0.61</u>
12			BERT _{UCSF} + LSTM + ST _{UCSF}		0.65	0.56	0.60
13			Shared task mean/median (10 teams)		0.68	0.58	0.62/0.70
14	Phase C (domain transfer)	MIMIC _{train+dev} (1504 docs.)	UW _{test} (518 docs.)	LSTM	0.86	0.66	0.75
15 [†]		+ UW _{train+dev} (2010 docs.)	*LSTM + fT _{MIMIC@UCSF}		0.85	0.70	0.77
16			LSTM + FLF _{UW}		0.85	0.68	0.76
17 [†]			*LSTM + fT _{MIMIC@UCSF} + FLF _{UW}		0.88	0.68	0.77
18 [†]			*LSTM + fT _{MIMIC@UCSF} + LW _{x10}		0.87	0.71	0.78
19			BERT _{UCSF} + LSTM		0.83	0.75	0.78
20			BERT _{UCSF} + LSTM + FLF _{UW}		0.84	0.74	0.79
21			BERT _{UCSF} + LSTM + LW _{x10}		0.82	0.75	0.78
22			Shared task mean/median (10 teams)		0.70	0.61	0.64/0.73

Association between SDoH and DNR/DNI status during hospital encounter

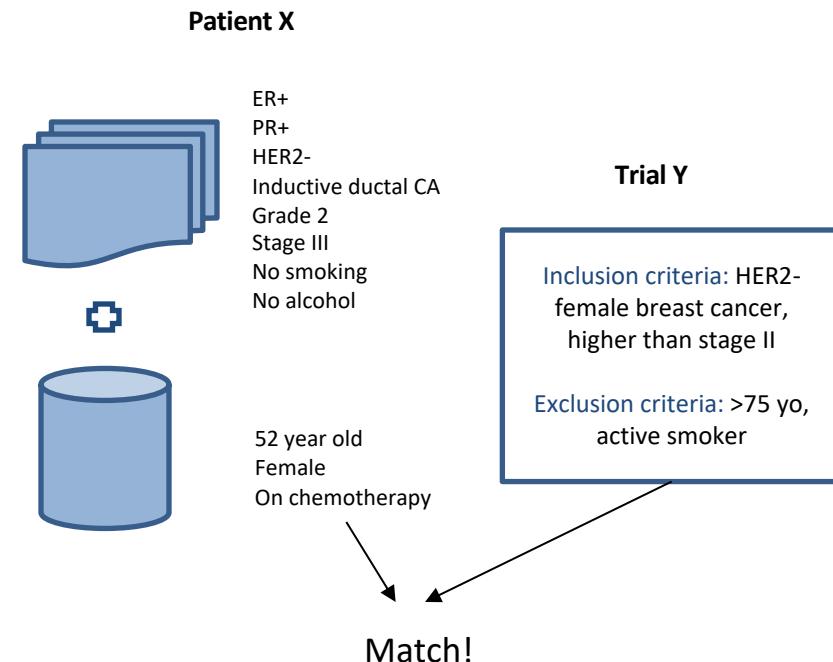
- For employment status, drug use and tobacco use, the point estimates from the manual labels differ meaningfully from those obtained with NLP models.
- No substantial differences on varying the number of examples either for transfer learning or model retraining.
- We find a negative association between being employed and the code status. However, the confidence intervals are very wide due to small sample size.



Promising directions for future research

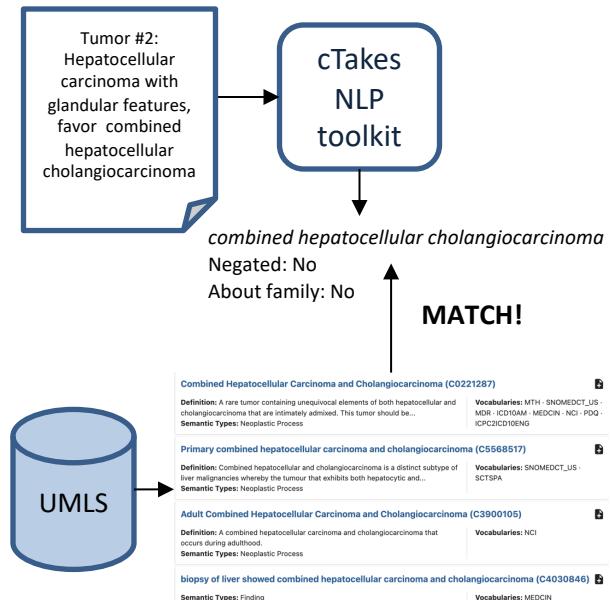
Matching patients to clinical trials

- Automated phenotyping to identify patient cohorts of interest
- Automatically matching patient phenotypes to trial inclusion and exclusion criteria
- Manual review of matched patient cohorts
- Expansion of patient cohorts if required by patient similarity analysis



Developing control arm for single-arm trials

- Automated phenotyping to identify patient cohorts of interest
- Monitoring real-world treatment administration and outcome
- Supplementing trial participant data with real-world observational data
- Particularly useful for rare diseases where patient recruitment is difficult
 - NLP-based phenotyping is superior to structured data-based phenotyping for diseases that do not have highly specific codes, for example, subtypes of liver cancer.
 - Lexical matching and clinical (UMLS) concept-based matching algorithms are often successful with a fraction of effort



Causal inference from textual data

- Inferring causality from observational data instead of just associations
 - Use of patient cohort matching, propensity score computations, and average treatment effect estimation
- Augmenting text-based features with structured data-based features for such inference
 - More complete data, since it can include fine-grained history from out-of-center encounters

Acknowledgements

Atul Butte, MD, PhD

Balu Bhasuran, PhD
James Buchanan, PharmD
Walter Daelemans, PhD
Artuur Leeuwenberg, PhD
Dana Ludwig, MD
Divneet Mandair, MD
Mahalakshmi Prakala
Yuwei Quan
Rebecca Racz, PharmD
Vivek Rudrapatna, MD, PhD
Anna Silverman, MD
Shenghuan Sun, BS
Simon Suster, PhD
Ahmed Wali
Yan-Ning Yu
Travis Zack, MD, PhD
Zhiwei Zheng

Oheneewaa Ahima, MD
Artur Belov, PhD
Lauren Choi, PharmD
Nadia Habal, MD
Samer El-Kamary, MD, MPH
Yan Li, B.Pharm, PhD
Qi Liu, PhD
Jawahar Tiwari, PhD

UCSF Information Commons
Wynton HPC team