A brief and rapid journal club of recent Clinical NLP work

Russ B. Altman, MD, PhD
Stanford University
UCSF-Stanford FDA CERSI

June 15, 2017

Method for this review

- Searched PubMED for (clinical…) AND (NLP…) AND (computational OR informatics…)
- Limited to 2016 and 2017
- ~200 articles curated down to 11 primary literature, 4 reviews, 4 “shout outs”

Useful Recent Reviews (methods & applications)

"Natural Language Processing Technologies in Radiology Research and Clinical Applications" (Cai et al, Radiographics) 26761536

"Natural Language Processing in Radiology: A Systematic Review" (Pons et al, Radiology) 27089187

"Natural Language Processing in Oncology" (Yim et al, JAMA Oncology) 27124593

"Extracting information from the text of electronic medical records to improve case detection: a systematic review" (Ford et al, JAMIA) 26911811

Table 3: Median accuracy by algorithm type and condition

	No. of Studies	Sensitivity (Recall)	Specificity	PPV (Precision)	Negative predictive value	F measure	AUROC
Algorithm type							
Single algorithm for NLP and case detection	15	96.2	97.4	85.35	96.6	49	–
Rule-based secondary case detection algorithm	20	91.2	95.45	77.5	98.95	97.57	94.4
Probabilistic secondary case detection algorithm (Logistic Regression; Bayesian; machine learning)	21	80	95	86	95.4	77	94
Condition							
Respiratory infections	11	92.9	95.45	54	99.9	–	95.85
Bowel disease	4	79.45	94.45	57.5	100	–	87.5
Inflammatory arthritis	5	70	96	93.7	–	–	94.4
Cancer	3	93	92.9	95	–	93.5	–
Diabetes	2	96.2	98	–	–	98.65	–
Obesity	5	48.4	–	76.3	–	49	–
Mental health	3	73.1	90	87.85	96.6	–	80
MRSA	2	99.2	99.4	97.9	–	99	–
Cardiovascular	7	82	96	84.7	93	74.85	92.9

"Computer-assisted expert case definition in electronic health records" (Walker et al, Int J Med Inf)

- Goal: Use NLP and other data sources for iterative cohort identification. Apply to acute liver dysfunc.
- Method: Start with seed group, iteratively sample, and update rules in response to expert curation.
- Result: Appy to 29K adults over 3 rounds of iteration, final definition sens 92%, PPV 79%.
- Conclusion: Can use NLP as part of a system to refine case definitions.

26725697

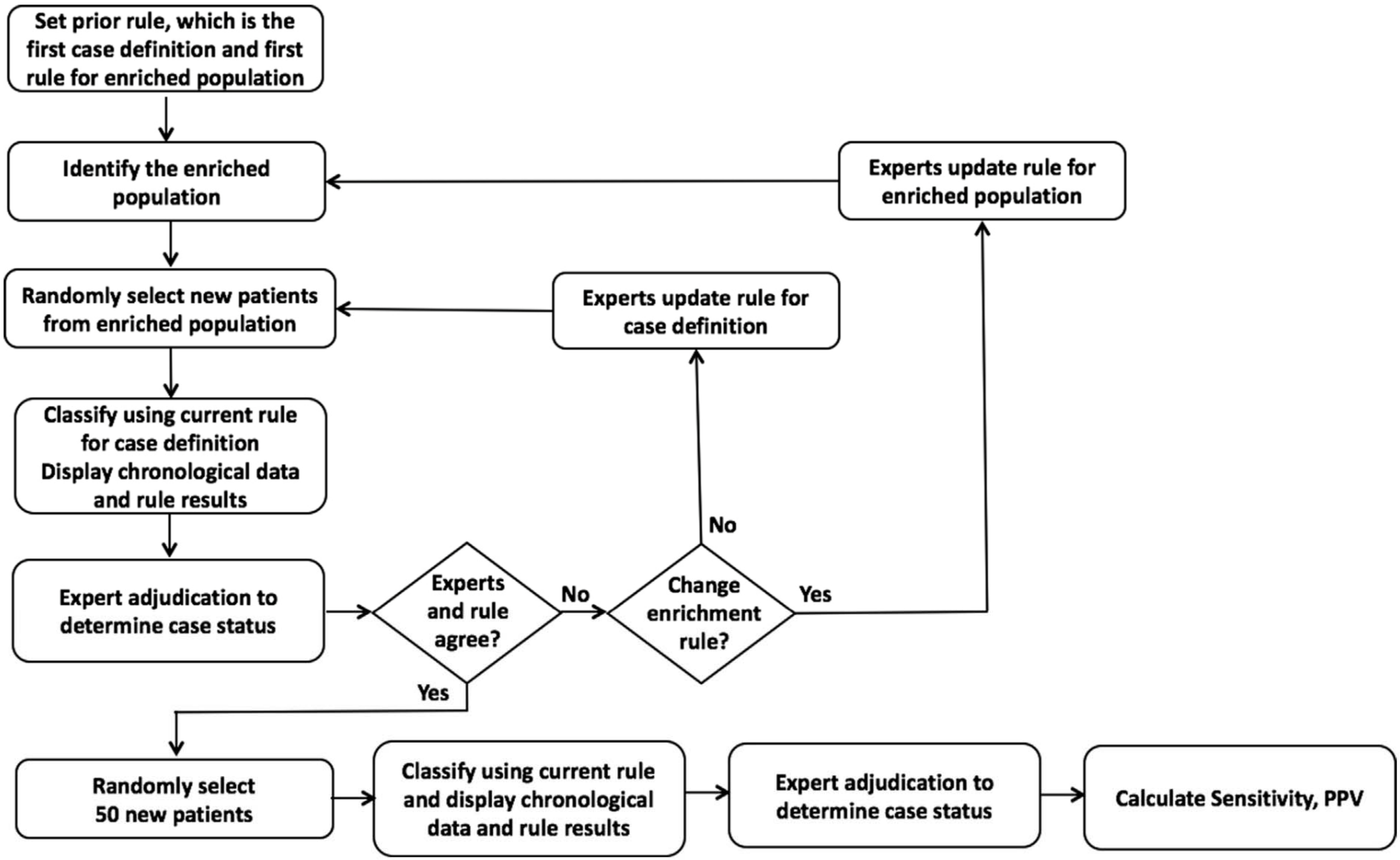


Fig. 1. Steps for computer-assisted development of a case definition.

Visit Date or Result Date	Diagnosis(Dx), Prescription(Rx), Lab, Procedure(PROC), Observations(OBS), Hospitalization(Hosp.), ER Visit, Dr. Notes(NLP)
31OCT2010	**Dx:** NAUSEA ALONE(78702) NONSPEC ELEV LEVEL TRANSAMINASE/LDH(7904) OTH NONSPEC ABN SERUM ENZYM LEVLS(7905) **Lab:** Albumin=4.2g/dl Alkaline phosphatase (ALP) (ALKP)=172u/l Alkaline phosphatase (ALP) (ALKP)=186u/l Bilirubin.direct=.1mg/dl Blood urea nitrogen=16mg/dl CO2.total=28mmol/l Calcium.total=9.1mg/dl Chloride=105mmol/l Glucose.random=110mg/dl Hepatitis B surface antigen (HBsAg).qualitative=non-reactive Hepatitis C antibody (HCAb) (anti-HCV).qualitative=negative O2 saturation.oximetry=95% Potassium=3.1mmol/l Protein.total=6.3g/dl Sodium (NA)=140mmol/l eGFR=89 **NLP:** **ALT--ALT(ELEVATED)** **NAUSEA--NAUSEA()**
01NOV2010	**Dx:** ABDOMINAL PAIN RIGHT UPPER QUADRANT(78901)
02NOV2010	**Lab:** Creatinine=1mg/dl **Bilirubin.total=.4mg/dl** **Alanine aminotransferase (ALT)=300u/l** **Aspartate aminotransferase (AST)=152u/l**
03NOV2010	**Dx:** NONSPEC ELEV LEVEL TRANSAMINASE/LDH(7904) **Lab:** O2 saturation.oximetry=95% **NLP:** **ALT--ALT(ELEVATED)** **ABDOMEN--PAIN(ABDOMENOCCASIONAL)** **ABDOMEN--PAIN(ABDOMENSOME;OCCASIONAL;CRAMPY)** **ABDOMEN--PAIN(ABDOMENWORSENING)** **OBS:** BMI=21.5 DBP=76 HR=72 SBP=122 WT=64.9

Fig. 2. Sample formatted patient data.

Table 3
Definition for identifying ALD cases after three revisions.

Inclusion criteria – *any of the following*

1. *Diagnostic codes*. Presence of an ICD-9 code corresponding to "Acute and subacute necrosis of the liver," "Other disorders of liver," or "Jaundice, unspecified, not of newborn";
2. *High transaminases*. AST or ALT values of at least three times the upper limit of normal (≥3xULN), provided that there had been a preceding value less than two times the upper limit of normal;
3. *Transaminases and bilirubin*. Either AST or ALT ≥2xULN and total bilirubin ≥3xULN or AST/ALT ≥3xULN and bilirubin ≥2xULN.

Onset date

The first date an inclusion criterion was met, or up to a week earlier if there were nonspecific hepatic ICD9 or NLP terms consistent with the onset of acute liver disease.

Exclusion criteria – *any of the following*

Related to onset

1. *Bile duct obstruction* codes within seven days of onset.
2. *Pancreatitis* codes within 30 days of onset.
3. *Viral hepatitis* codes or NLP within 30 days of onset.
4. *Metastatic cancer* codes within 30 days of onset.
5. *Chemotherapy* codes within 30 days of onset.
6. Liver transplant before onset.

Related to underlying conditions

7. *Cholangitis* codes appearing at least two times in the record.
8. *Persistent transaminase elevation*. Multiple values of AST/ALT ≥3xULN, not contained within a 183-day window and with no AST/ALT values <2xULN intervening. Does not exclude when there is an AST/ALT ≥6xULN.
9. *Other chronic conditions*. Codes for sarcoidosis or other chronic nonalcoholic liver disease, plus NLP terms for cirrhosis.

Rule

Subjects with an inclusion criterion and no exclusion criterion were counted as cases of acute liver disease with onset as specified.

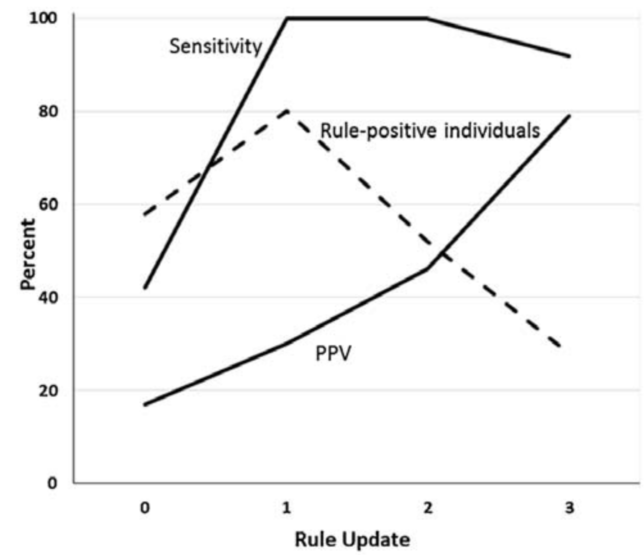
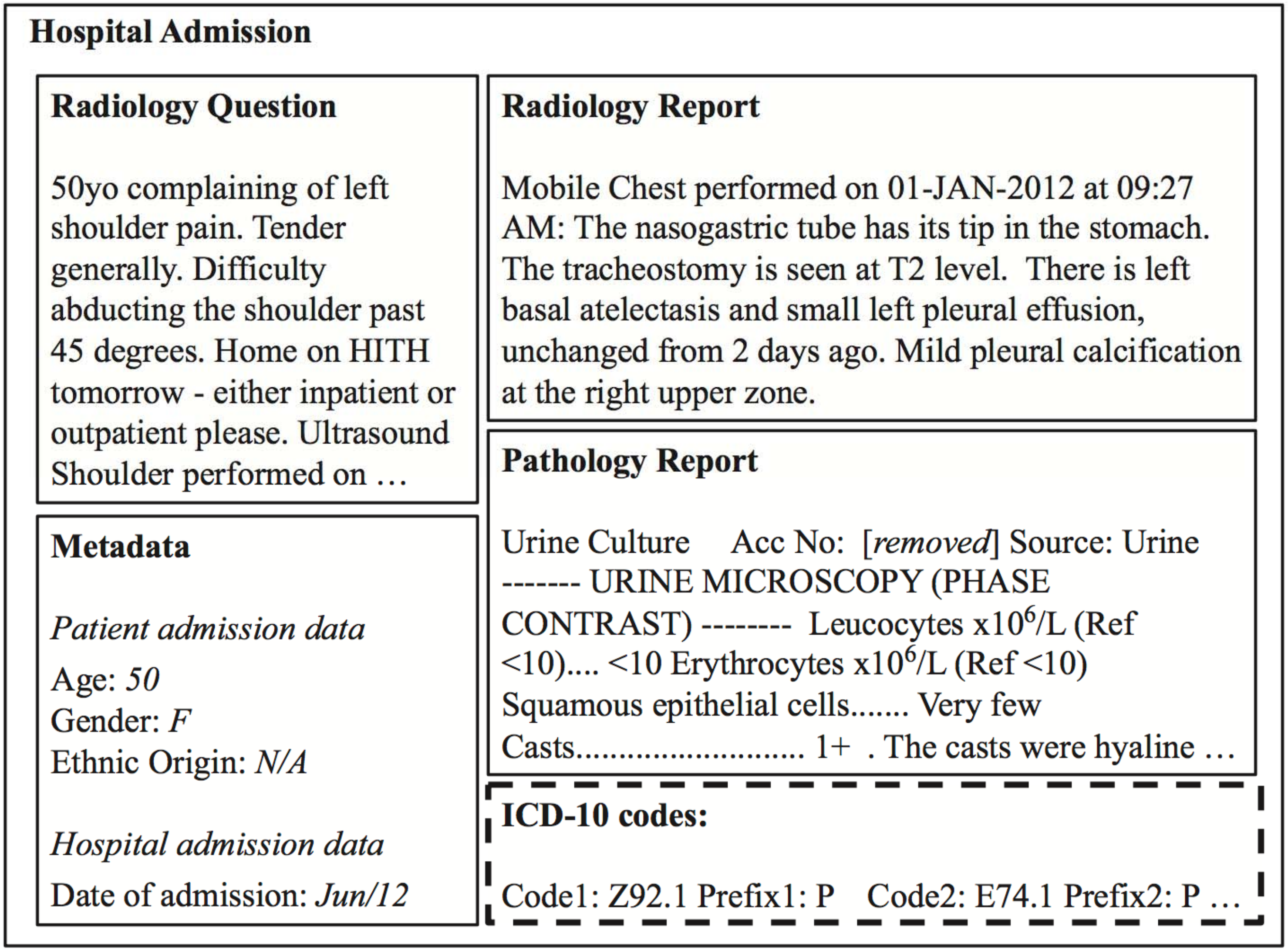


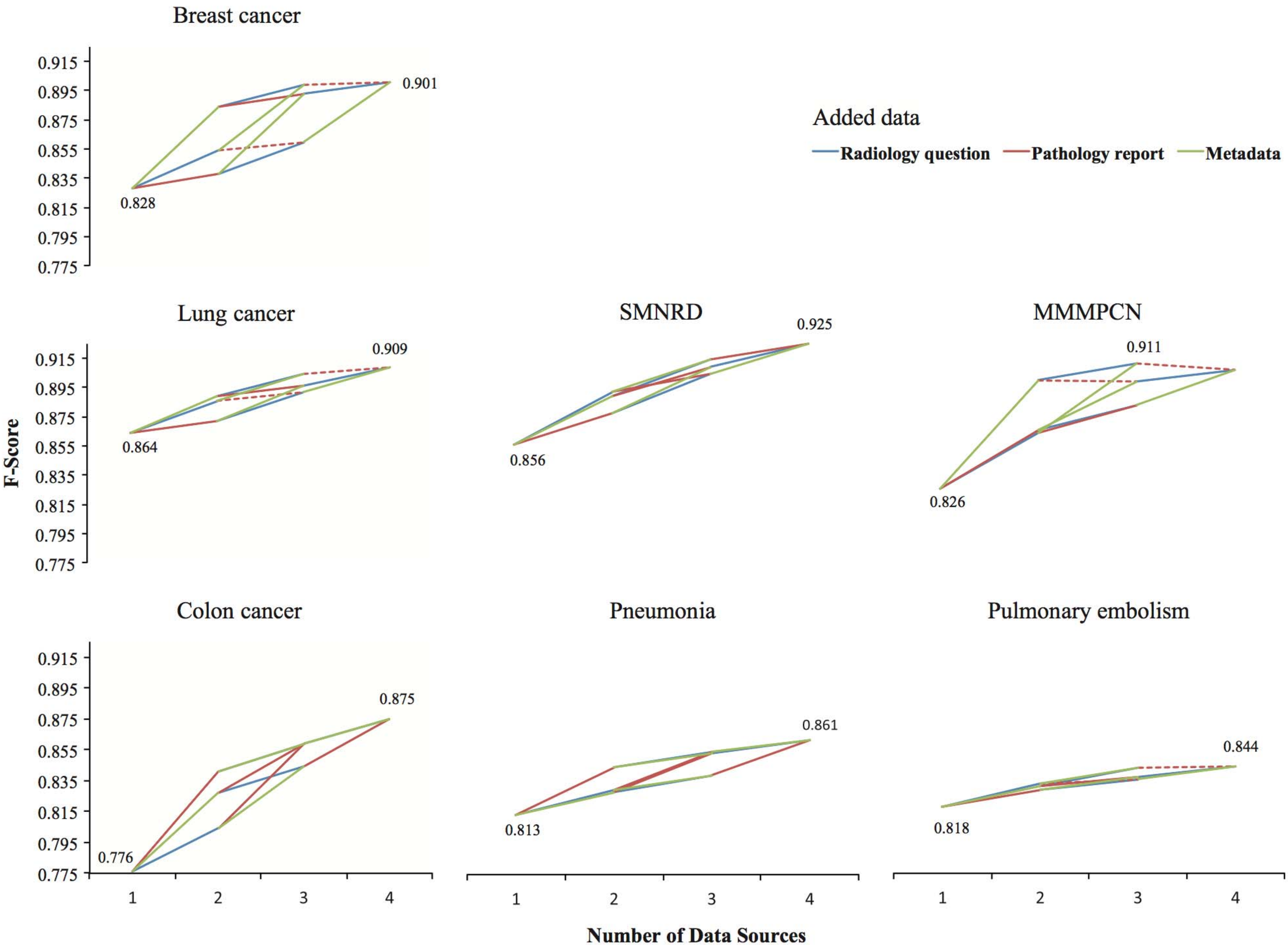
Fig. 3. Sensitivity, rule-positive percent and PPV with sequential rule updates. All rules applied to the fourth (test) data set.

“Text mining electronic hospital records to automatically classify admissions against disease: Measuring the impact of linking data sources” (Kocbek et al, J Biomed Inf)

- Goal: Examine effect of multiple data sources at recognizing 8 key diseases
- Method: SVM with radiology, pathology, patient and hospital admission data. “NLP” using MetaMap/UMLS.
- Result: Radiology, patient, hospital data most useful for detecting diseases.
- Conclusion: Linking data sources improves overall performance

27742349


Hospital Admission
Radiology Question
50yo complaining of left shoulder pain. Tender generally. Difficulty abducting the shoulder past 45 degrees. Home on HITH tomorrow - either inpatient or outpatient please. Ultrasound Shoulder performed on …
Radiology Report
Mobile Chest performed on 01-JAN-2012 at 09:27 AM: The nasogastric tube has its tip in the stomach. The tracheostomy is seen at T2 level. There is left basal atelectasis and small left pleural effusion, unchanged from 2 days ago. Mild pleural calcification at the right upper zone.
Pathology Report
Urine Culture Acc No: [*removed*] Source: Urine ------- URINE MICROSCOPY (PHASE CONTRAST) -------- Leucocytes x10^6/L (Ref <10).... <10 Erythrocytes x10^6/L (Ref <10) Squamous epithelial cells....... Very few Casts......................... 1+ . The casts were hyaline …
Metadata
Patient admission data
Age: *50*
Gender: *F*
Ethnic Origin: *N/A*
Hospital admission data
Date of admission: *Jun/12*
ICD-10 codes:
Code1: Z92.1 Prefix1: P Code2: E74.1 Prefix2: P …



Breast cancer
0.901
0.828
Added data
Radiology question
Pathology report
Metadata
Lung cancer
0.909
0.864
SMNRD
0.925
0.856
MMMPCN
0.911
0.826
F-Score
Colon cancer
0.875
0.776
Pneumonia
0.861
0.813
Pulmonary embolism
0.844
0.818
0.915
0.895
0.875
0.855
0.835
0.815
0.795
0.775
1
2
3
4
Number of Data Sources

Radiology

“Using automatically extracted information from mammography reports for decision-support” (Bozcurt et al, J Biomed Inf)

- Goal: Combine NLP extraction with Bayesian decision support for breast cancer dx.
- Method: NLP system generates input to Bayes Network. Predict prob(malig) & BI-RADS assessment.
- Result: 98% performance on BI-RADS. 95% concordance with reference on malignancy
- Conclusion: NLP can generate features for accurate patient classification.

26911811

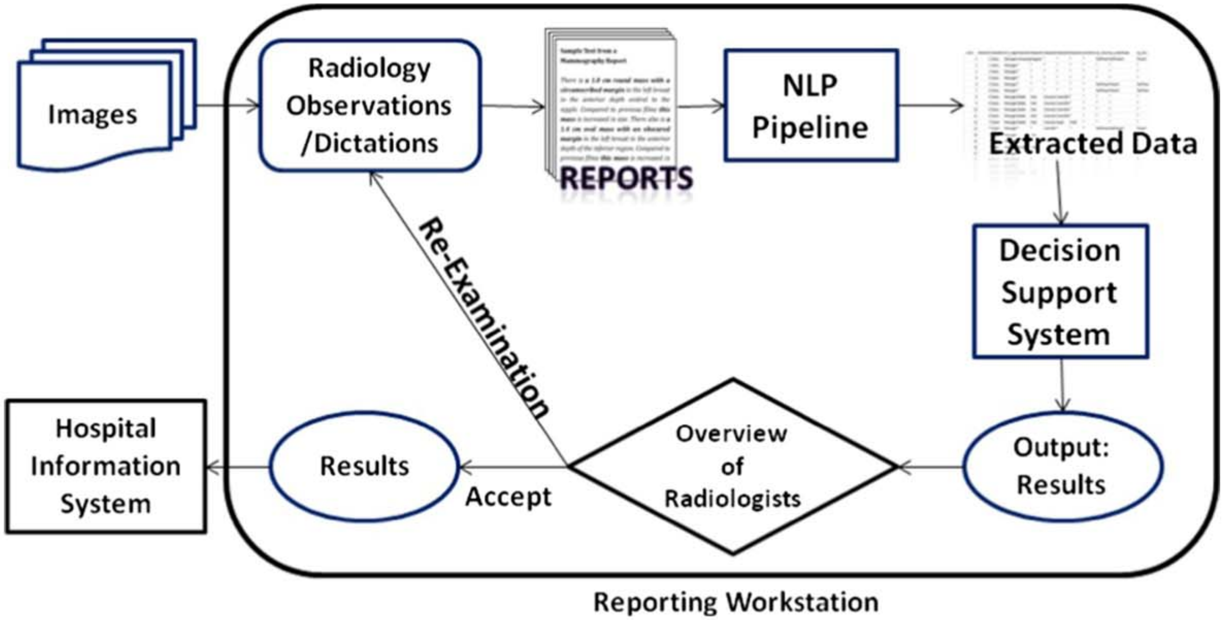
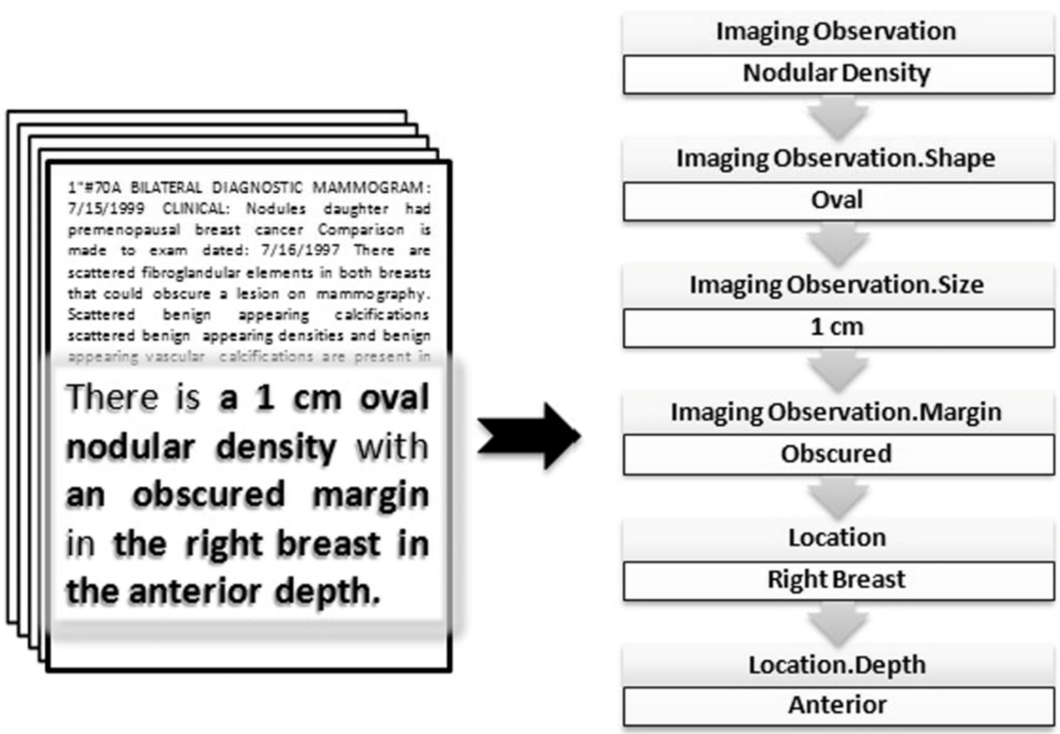


Fig. 1. Decision support tools developed and their relationship to the radiology workflow.



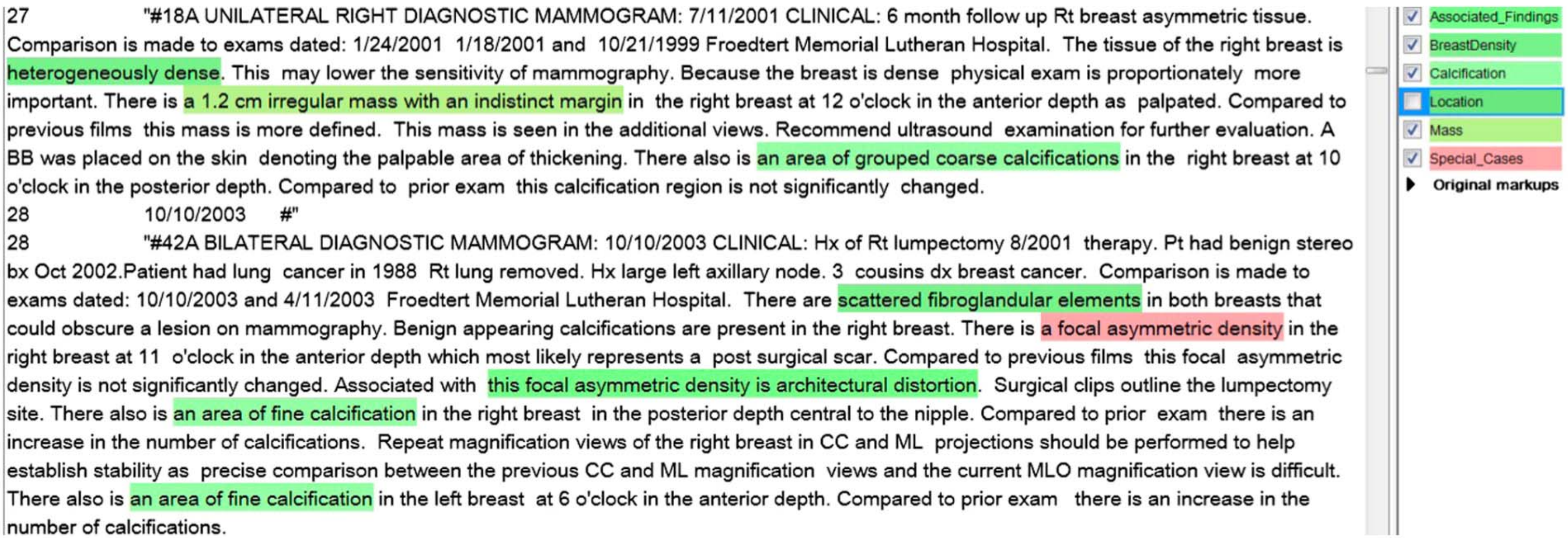


Fig. 2b. Output from NLP system on GATE NLP GUI.

“Information extraction from multi-institutional radiology reports” (Hassanpour et al, AI in Med)

- Goal: Extract key features from radiology reports to inform clinical QC & research.
- Method: Discriminative sequence classifiers for NER. Applied to 3 healthcare systems. Used information model of radiology concepts.
- Result: Extract concepts with F1=85%
- Conclusion: Effective method to annotate and extract clinical information from free text for clinical assistance and research.

26481140

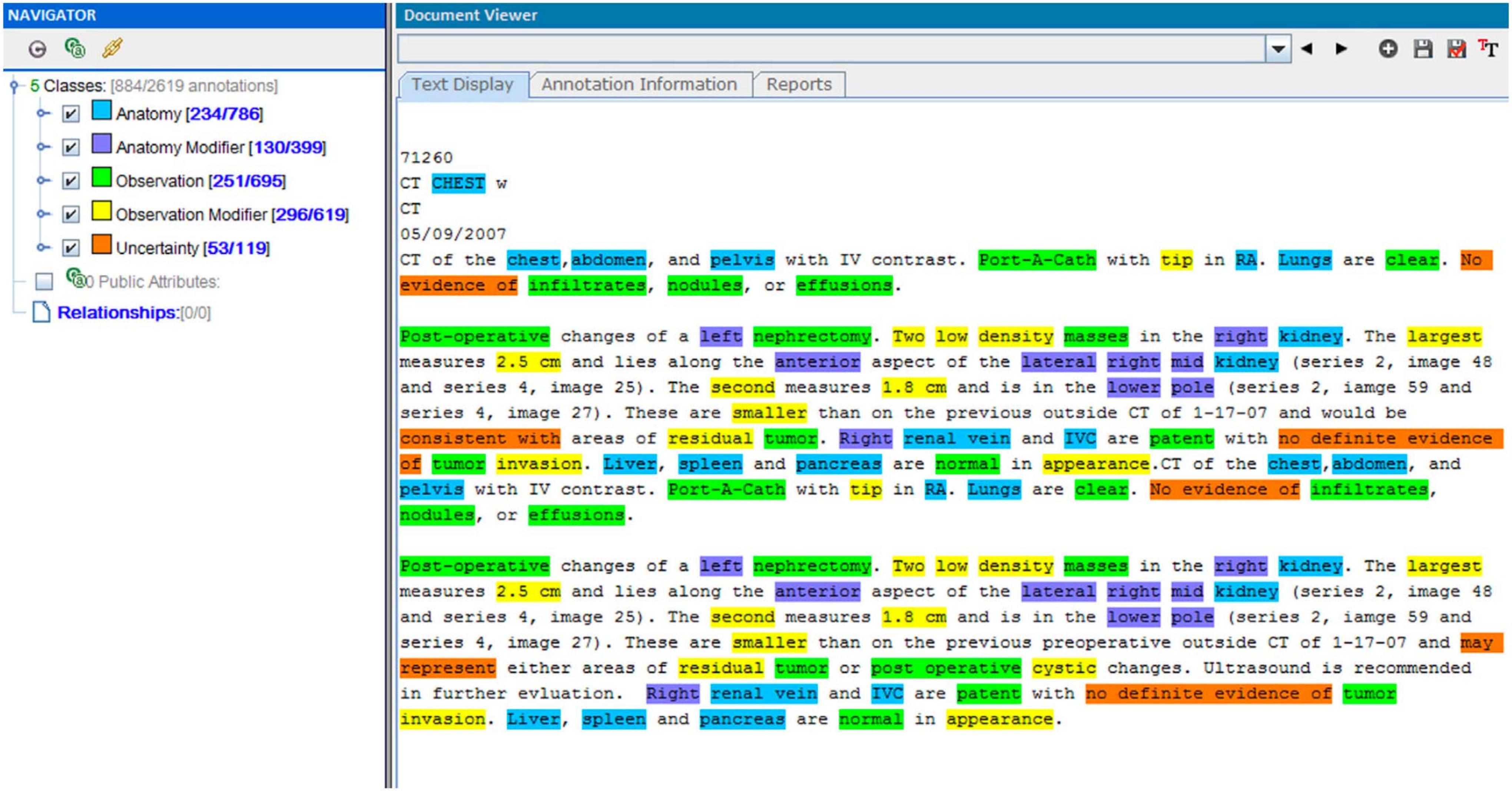


Fig. 1. A sample manually annotated radiology report in eHOST.

“Characterization of Change and Significance for Clinical Findings in Radiology Reports Through Natural Language Processing” (Hassanpour et al, J Dig Imag)

- Goal: Identify significant changes across radiographs over time.
- Method: Rules and machine learning to extract NLP features. Trained for change and significance.
- Result: Change F1 = 95%, Significance F1 = 75%
- Conclusion: Automated use of NLP to detect significant radiographic changes, help MDs.

28050714

Change class	Rule
New or worse	Presence of “new”, “increase”, “develop”, “progress”, and “more” in positive context
Unchanged	Presence of “new”, “increase”, “develop”, “progress”, and “more” in negative context
	Presence of “change” in negative context
	Presence of “stable”, “remain”, and “persist”
Improved	Presence of “improve” and “decrease”
Indeterminate	Absence of other criteria

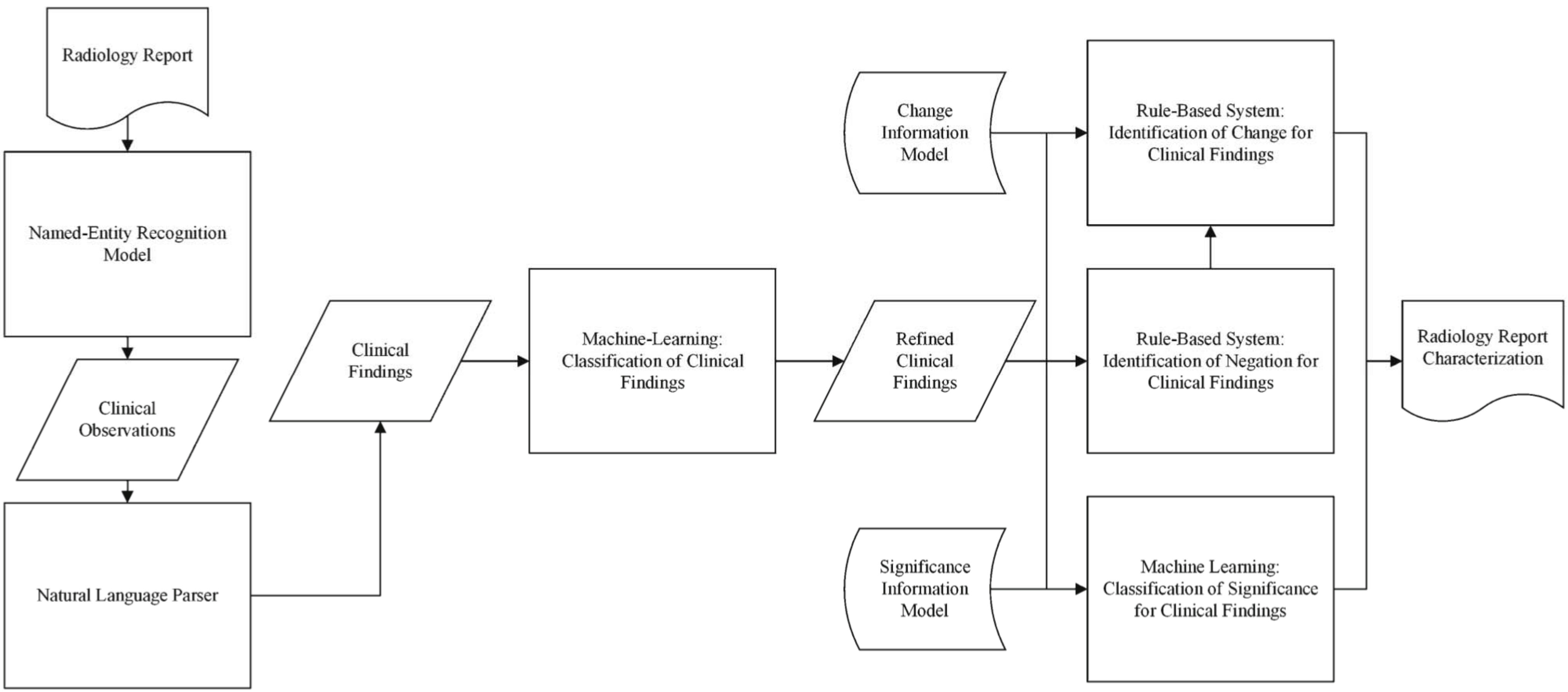


Fig. 1 Overview of our methodology for characterization of change and significance for clinical findings in radiology reports

Specific Specialties

"Congestive heart failure information extraction framework for automated treatment performance measures assessment" (Meystre et al, JAMIA)

- Goal: Extract information about treatment performance for CHF from text.
- Method: Rules, dictionaries, ML to extract information about heart function (e.g. LVEF).
- Result: Medications, LVEF quantitative 91-98%. Reasons for CHF med use 40%/30% recall/prec.
- Conclusion: Can find key information about CHF in text, but some categories more difficult.

27413122

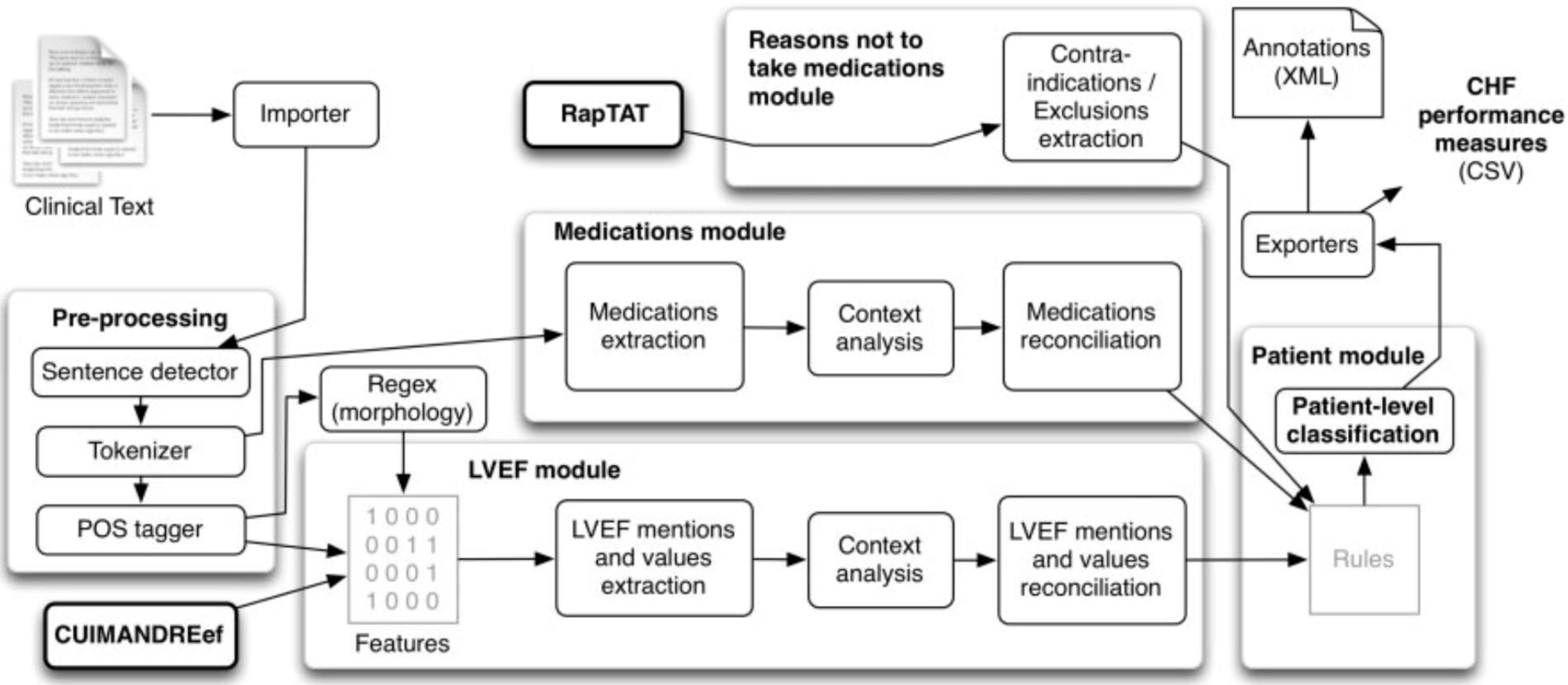


Figure 1. CHIEF general architecture.

Table 2. CHIEF information extraction results with exact matches (with 0.95 binomial exact confidence intervals)

Information extracted	N	Recall	Precision	F_1-measure
Mentions of LVEF	2276	0.978 (0.971–0.984)	0.986 (0.980–0.990)	0.982
LVEF quantitative values	2200	0.910 (0.897–0.921)	0.939 (0.928–0.949)	0.924
ACEI medications	2949	0.994 (0.990–0.996)	0.976 (0.970–0.981)	0.985
ARB medications	591	0.978 (0.963–0.988)	0.960 (0.941–0.974)	0.969
Reasons not to take ACEI/ARB	483	0.311 (0.270–0.354)	0.247 (0.213–0.283)	0.275
Overall (micro-average)	8499	0.928 (0.922–0.933)	0.917 (0.911–0.923)	0.922

Table 3. CHIEF information extraction results with partial matches (with 0.95 binomial exact confidence intervals)

Information extracted	N	Recall	Precision	F_1-measure
Mentions of LVEF	2276	0.986 (0.980–0.990)	0.994 (0.990–0.997)	0.990
LVEF quantitative values	2200	0.945 (0.934–0.954)	0.976 (0.968–0.982)	0.960
ACEI medications	2949	0.996 (0.993–0.998)	0.978 (0.972–0.983)	0.987
ARB medications	591	0.997 (0.988–1.000)	0.978 (0.963–0.988)	0.987
Reasons not to take ACEI/ARB	483	0.404 (0.360–0.449)	0.321 (0.284–0.359)	0.358
Overall (micro-average)	8499	0.946 (0.941–0.951)	0.935 (0.930–0.940)	0.940

“Natural language processing to extract symptoms of severe mental illness from clinical text: the Clinical Record Interactive Search Comprehensive Data Extraction (CRIS-CODE) project” (Jackson et al, BMJ Open)

- Goal: Use NLP to capture info for mental illness
- Method: Build & Apply rules to 23K discharge summaries. Attempted 50 key symptoms
- Result: For 46 symptoms, F1 = 88%. Successful symptoms from 87% with MI, 60% without.
- Conclusion: Symptoms of MI can be extracted.

28096249

Table 1 Symptom instance definitions

SMI concept	Keyword strings	Modifier strings	Lax or strict modifiers	SNOMED-CT (SCTID)†
Aggression	aggress*			61372001
Agitation	agitat*			106126000
Anhedonia	anhedon*			28669007
Apathy	apath*			20602000
Arousal	arous*			(none)
Blunted or flat affect	Affect	blunt*, flat*, restrict*	Optional	6140007/932006/39370001
Catalepsy	catalep*			247917007
Catatonic syndrome	catatoni*			247917007
Circumstantial speech	circumstan*			18343006
Deficient abstract thinking	Concrete			71573006
Delusions	delusion*			2073000
Derailment of speech	derail*			65135009
Diminished eye contact	eye contact			412786000
Disturbed sleep	Sleep	not, poor*, interrupt*, nightmare*, disturb*, inadequat*, disorder*, prevent*, stop*, problem*, difficult*, reduced*, less*, impair*, erratic*, unable*, worse*, depriv*	Optional	26677001

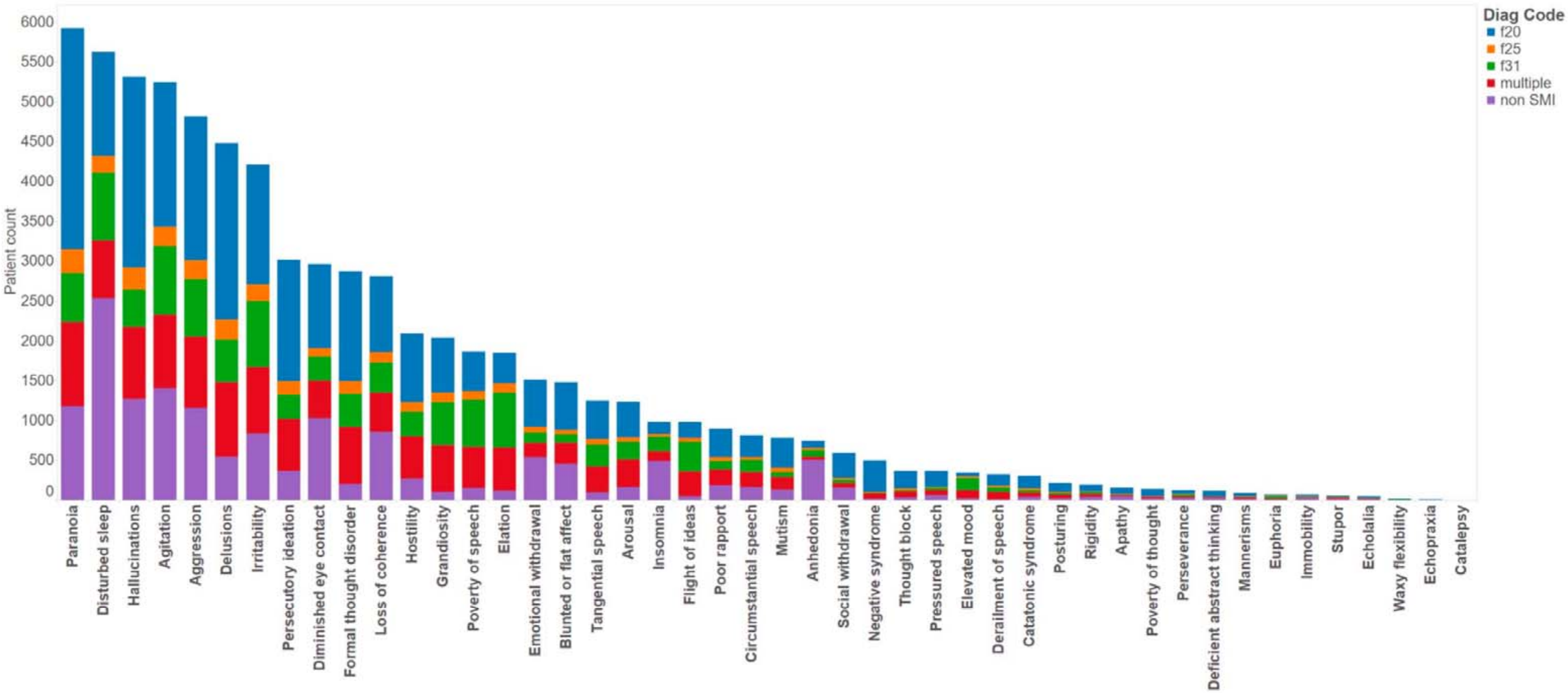


Figure 1 Distribution of symptoms by SMI ICD diagnosis. ICD, International Classification of Diseases; SMI, severe mental illness.

clinical notes of patients with heart diseases: Developing and validating a natural language processing application” (Topaz et al, Int J Nurs Stud)

- Goal: Extract information about patient wounds. Only 46% of charts have structured wound info.
- Method: MTERMS engine for NLP. Built information model using terminologies + keywords.
- Result: 101 notes. F-measure 93%. Best = wound treatment, worst = wound size.
- Conclusion: Free text can supplement poorly coded data with extraction of key wound

27668855

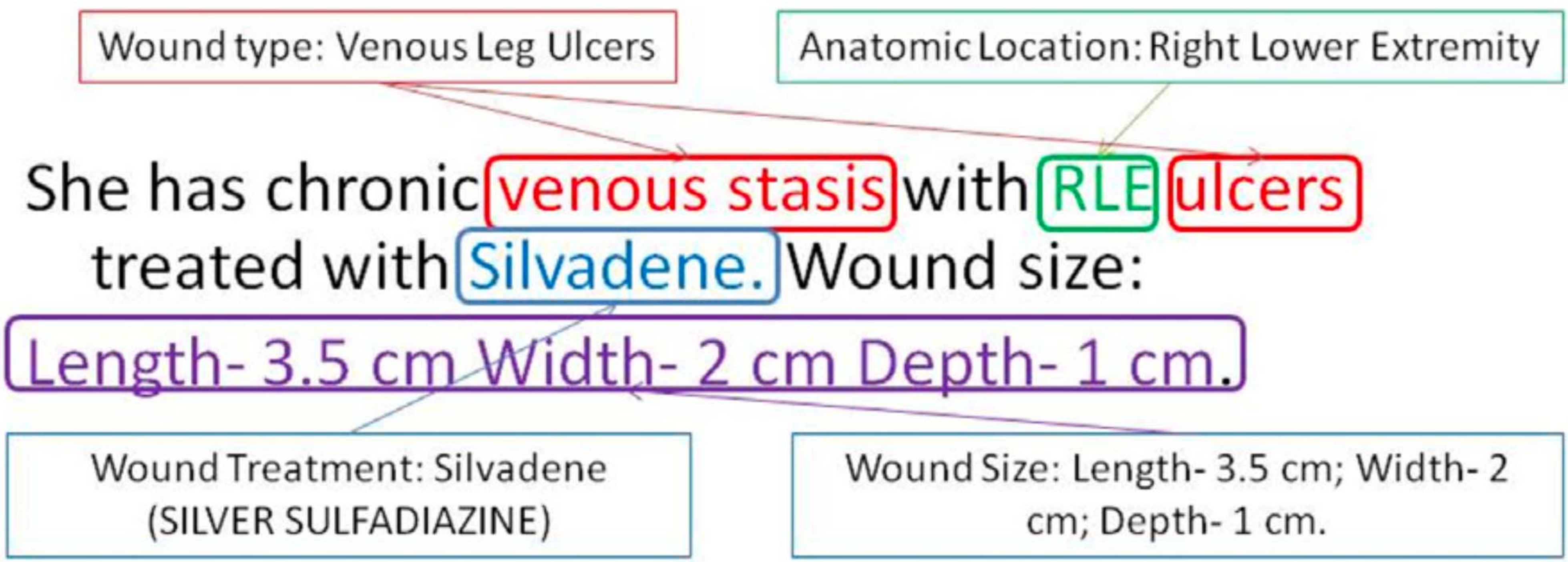


Fig. 2. Wound information annotation example.

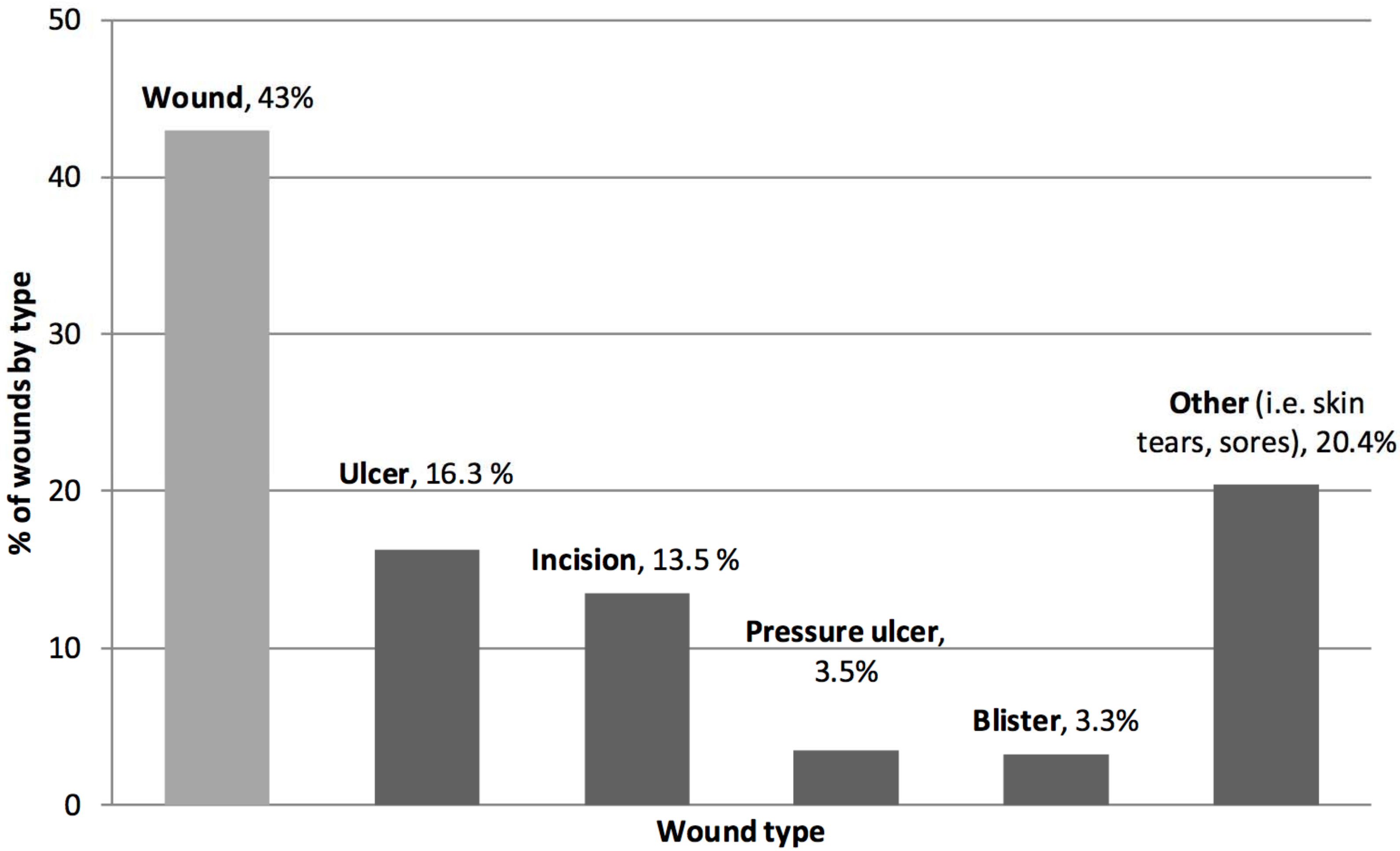


Fig. 3. Distribution of wound type information in the clinical notes.

Other settings

“Identification of Nonresponse to Treatment Using Narrative Data in an Electronic Health Record Inflammatory Bowel Disease Cohort” (Ananthakrishnan et al, Inflam Bowel Dis)

- Goal: NLP of charts to characterize response to IBD treatment (antibodies to TNF-alpha)
- Method: Regression on narrative text, coded complications, procedures & med features for 3 classes: response, partial, non-response
- Result: 33% correlation with MD grading, 18% nonresponse, 21% partial, 56% complete.
- Conclusion: Promising initial results for drug response

26332313

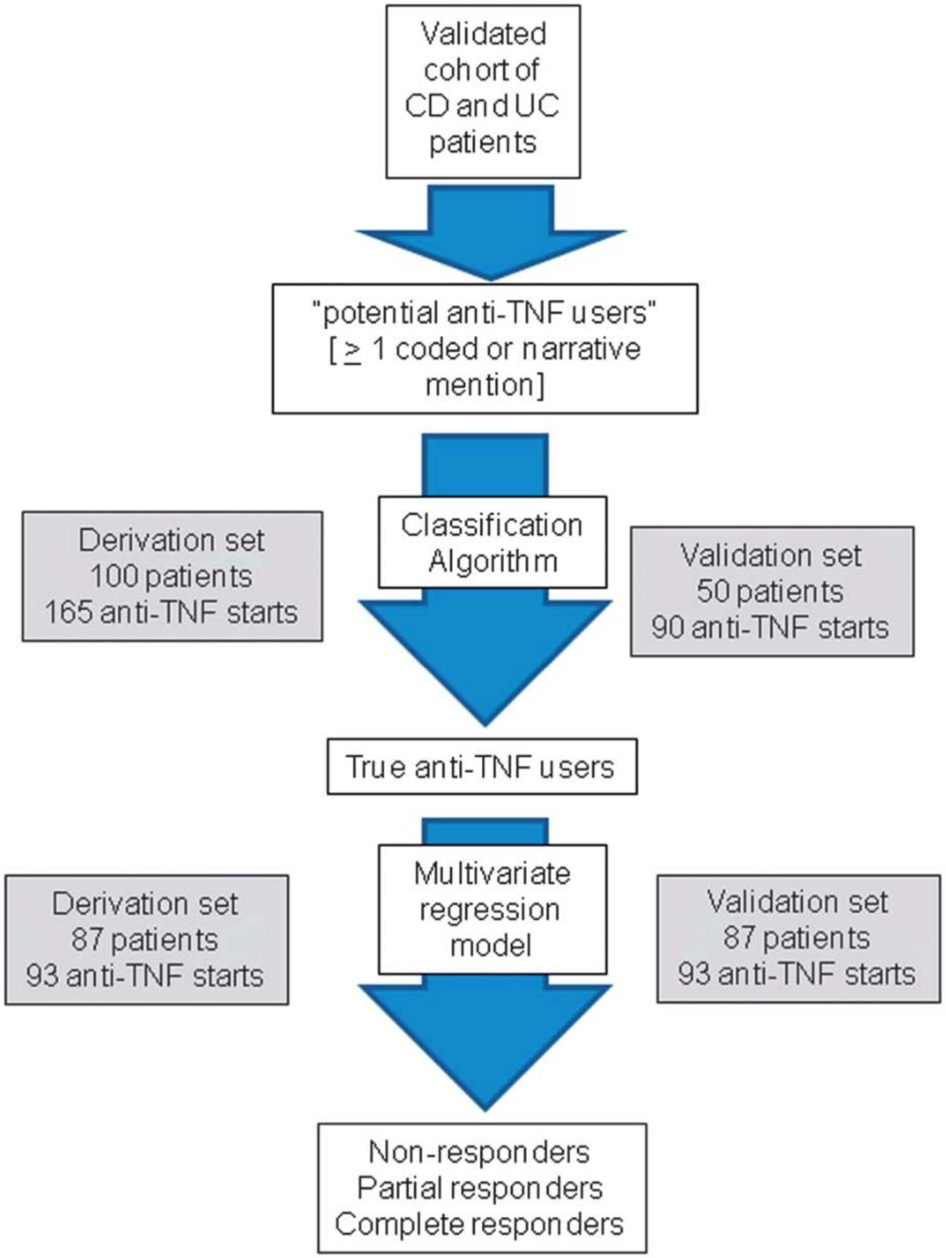


FIGURE 1. Flowchart representing study procedure to identify non-response to treatment in the EHR.

TABLE 2. Univariate Analysis of Predictors of Nonresponse to Anti-TNF Biologic Therapy in IBD

Term	Odds Ratio for Nonresponse
Pain	1.01 (1.00–1.02)
Diarrhea	1.09 (1.03–1.15)
Relapse	1.12 (0.59–2.11)
Abdomen	1.01 (1.00–1.02)
Bleeding	1.01 (0.92–1.10)
Tenesmus	1.01 (0.50–2.05)
Colonoscopy	1.03 (0.99–1.08)
Fatigue	1.18 (1.04–1.34)
Absence of pain	0.92 (0.65–1.29)
Remission	1.00 (0.89–1.12)

“Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database” (Karystianis et al, BMC Med Inf)

- Goal: Extract detailed medication dose information from free text.
- Method: Rules to extract applied to primary care data
- Result: On 220 free text directions, 91% accurate a prescription level, 97% across individual attributes.
- Conclusion: Good ability to extract detailed dosing information from free text.

26860263

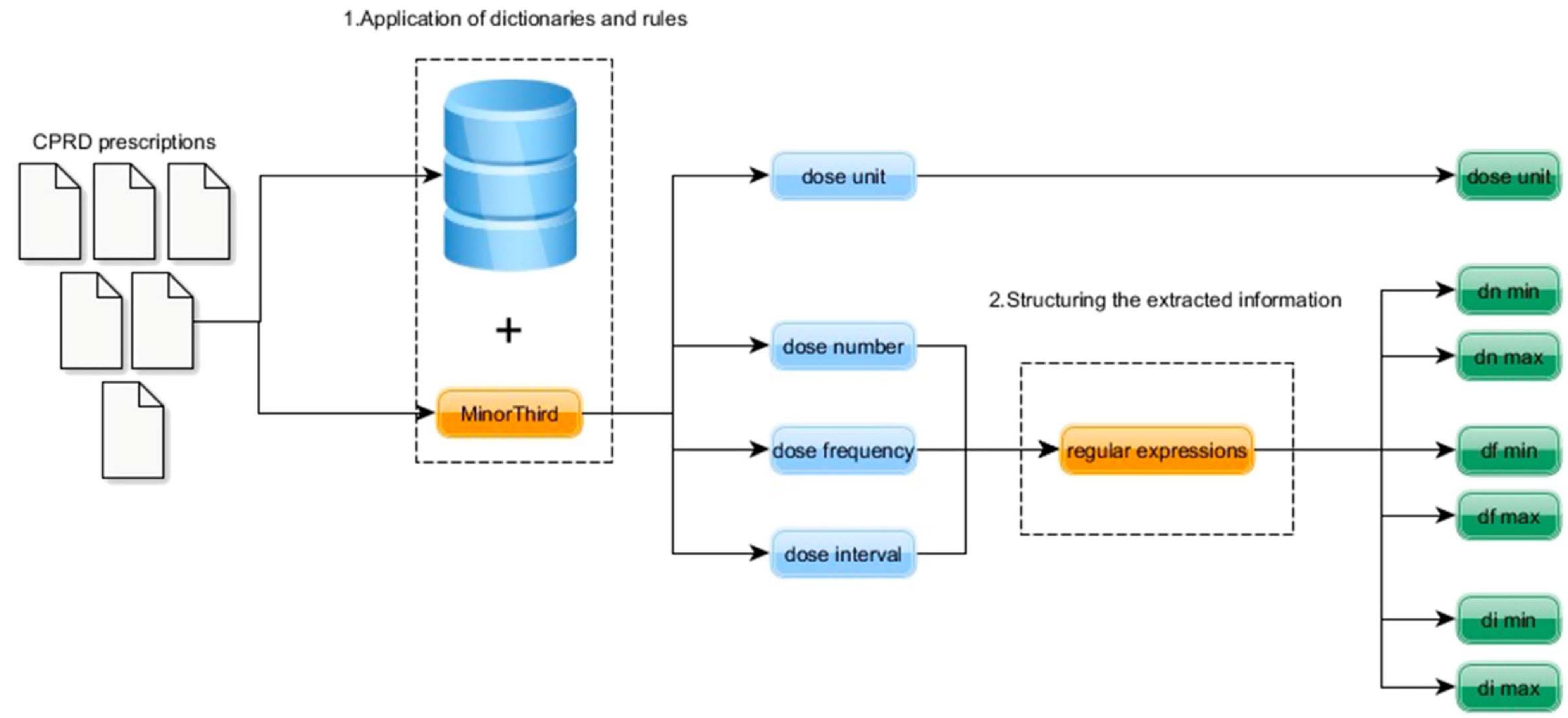


Fig. 1 The two-step approach for the extraction of structured dose information from CPRD prescription instructions

Table 1 Examples of prescription instructions represented in our model

Prescription	dn_min	dn_max	df_min	df_max	di_min	di_max	dose unit
take 2 tablets 4 times a day	2	2	4	4	1	1	tablet
2 tabs qid	2	2	4	4	1	1	tablet
a half to one tablet to 2 three times a day when required	0.5	2	0	3	1	1	tablet
10 mg to be taken weekly	10	10	1	1	7	7	mg
2 with each meal	2	2	3	3	1	1	?
take 2.5 ml twice a day	2.5	2.5	2	2	1	1	ml
half a tablet twice a day when required	0.5	0.5	0	2	1	1	tablet
2 puffs 6 hrly prn	2	2	0	4	1	1	puff
1 to 3 every day	1	3	1	1	1	1	?
one or two to be taken every 4 to 6 hours	1	2	4	6	1	1	?
take as directed	1	?	?	?	1	?	-
apply as needed	1	1	0	?	1	?	-

dn_min is dose number (minimum), *dn_max* is dose number (maximum), *df_min* is dose frequency (minimum), *df_max* is dose frequency (maximum), *di_min* is dose interval (minimum), *di_max* is dose interval (maximum). Additional file 1: Table S1 contains examples of frequent Latin abbreviations

Table 2 Examples of rules for the recognition of dosage attributes in medication data

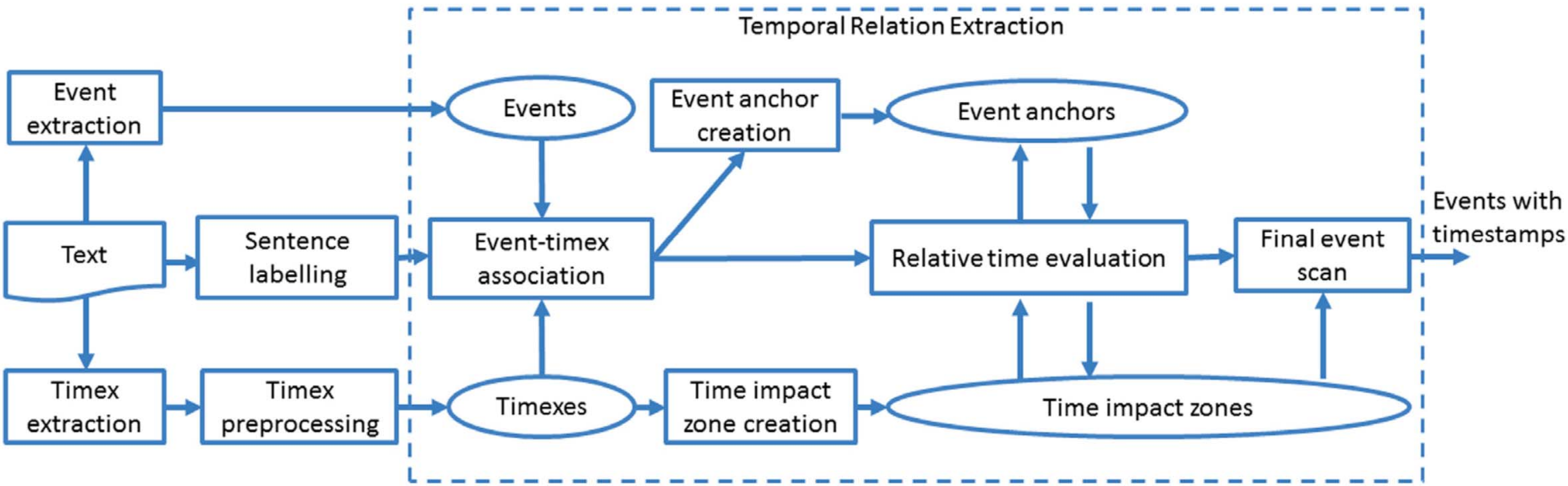
<table>
<tr><th>dosage attribute (number of rules)</th><th>examples</th><th colspan="5">identified span (in bold)</th></tr>
<tr><td rowspan="5">dose number
(149 rules)</td><td>rule →</td><td>a(verb)</td><td>[@number]</td><td>a(DoseUnit)</td><td>eq('each')</td><td>@period</td></tr>
<tr><td>take two capsules each morning</td><td>Take</td><td>two</td><td>capsules</td><td>each</td><td>morning</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td>rule →</td><td colspan="2">a(verb)</td><td colspan="2">[@number]</td><td>a(timeUnitLy)</td></tr>
<tr><td>take one daily</td><td colspan="2">take</td><td colspan="2">one</td><td>daily</td></tr>
<tr><td colspan="7"></td></tr>
<tr><td rowspan="5">dose frequency
(90 rules)</td><td>rule →</td><td colspan="2">a(verb)</td><td colspan="2">[@number]</td><td>eq('times')</td></tr>
<tr><td>per day apply 3-4 times</td><td colspan="2">Apply</td><td colspan="2">3-4</td><td>times</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td>rule →</td><td colspan="2">a(verb)</td><td>@number</td><td>a(DoseUnit)</td><td>[@perTimeUnit]</td></tr>
<tr><td>inhale 2 puffs three times a day</td><td colspan="2">inhale</td><td>2</td><td>puffs</td><td>three times a day</td></tr>
</table>

"A new algorithmic approach for the extraction of temporal associations from clinical narratives with an application to medical product safety surveillance reports" (Wang et al, J Biomed Inf)

- Goal: Tag and extract temporal information from narratives, associate with related events.
- Method: "Shallow" syntactic information extracted with rules. Built model of temporal statements.
- Result: 86-88% F-measure on 140 FAERS/VAERS. Correct on 69% of event relations in i2b2 test set
- Conclusion: Temporal information models can extract useful temporal data from narrative text.

27321529

- Date – if a time string does NOT contain time intervals such as “day”, “month”, “year”, “hours”, etc.
- Age – if a time string is followed by “old”, “of age”, “y/o”, or “yo”.
- Relative– if the time string:
  - is adjacent to relative signal words, such as “before”, “after”, “prior”, “later”, “earlier”, “post”, “ago”, “next”, and “following”.
  - contains words like “same”, “time”, “morning”, “evening”, as well as weekdays such as: “Sunday”, and “Monday”.
- Duration – if the time string follows words like “for”, “over”, “lasting”, and “within”.
- Frequency – if the time string follows words like “every” or “per”.



Temporal Relation Extraction
Event extraction
Events
Event anchor creation
Event anchors
Text
Sentence labelling
Event-timex association
Relative time evaluation
Final event scan
Events with timestamps
Timex extraction
Timex preprocessing
Timexes
Time impact zone creation
Time impact zones

Shout Outs:

"Unsupervised entity and relation extraction from clinical records in Italian" (Alicante et al, Comp in Bio Med)

26851833

"Mining Clinicians' Electronic Documentation to Identify Heart Failure Patients with Ineffective Self-Management: A Pilot Text-Mining Study " (Topaz et al, Nursing Inf)

26851833

"Normalizing clinical terms using learned edit distance patterns" (Kate, JAMIA)

26232443

"Clinical records anonymisation and text extraction (CRATE): an open-source software system" (Cardinal, BMC Med Inf Dec Mak)

28441940

Thanks.

russ.altman@stanford.edu