Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Journal Club : Apr 19, 2024

Toward structuring real-world data: Deep learning for extracting oncology information from clinical text with patient-level supervision

From Patterns 4, 100726, April 14, 2023

Presenter : Ekapob Sangariyavanich ,MD

Wisdom of the Land

Introduction

- In the US, cancer is a reportable disease, and cancer centers are required to curate patient information per national accreditation and clinical quality requirements.
- Manual curation clinical text, such as pathology assessments, radiology assessments, and clinical progress notes to structure such text is expensive and hard to scale.
- Natural language processing (NLP) can help accelerate manual curation.

- Previous studies

  - Gao et al. and Kehl et al. restricted classification to individual pathology reports and exclude tumors associated with multiple reports.
  - Percha et al. focus on classifying individual sentences for breast cancer surgery information.
- Such methods are not applicable to the prevalent cases where information is scattered across multiple clinical documents.
- Information in a single document is insufficient, and additional context is required for identifying the correct diagnosis or staging information.

- We propose to bootstrap deep learning for structuring RWD by using readily available registry data. By matching registry entries with their corresponding EMR data.
- To the best of our knowledge, our study is the first to explore cross-document medical information extraction using registry derived, patient-level supervision to train deep NLP methods.

Methods

Data

- From 28 distinct cancer care centers across US states.
- Matching comprehensive EMR records (including all free-text clinical documents) and cancer registry records.
- Exclude: patients without a digitized pathology report within 30 days of diagnosis.
- Total of 135,107 patients spanning multiple US states between 2000 and 2020.

- Using patients in Oregon for the initial exploration (n = 39,064 (29%) of patients).
- We divide patients into 10 random folds.
  - Training and development: 6 folds (n = 23,438)
  - Test: 2 folds (n = 7,745)
  - Additional held-out test set: 2 folds (n = 7,881)
- Patients from Washington (n = 36,900), as well as the remaining states (n = 59,143) for further generalizability tests.

- Interested core cancer attributes:
  - Tumor site (330 classes)
  - Histology (556 classes)

  } ICD-O-3 ontology

  - Staging --- AJCC (T:0–4, in situ ; N:0 vs.1+ ; M: 0 vs 1)
- EMR: (concatenated chronologically)
  - Pathology report
  - Radiology reports
  - Operative notes.
- Metric: AUROC, AUPRC

Model comparison

- Ontology-aware rule-based system
- Context free : - BOW
  - OncoGLove
- Context : - BERT
  - PubMedBERT
  - OncoBERT

+

-CNN
-HAN/GRU

CNN: convolutional neural network
HAN: hierarchical attention network
GRU : gated recurrent unit

Multiclass classifier

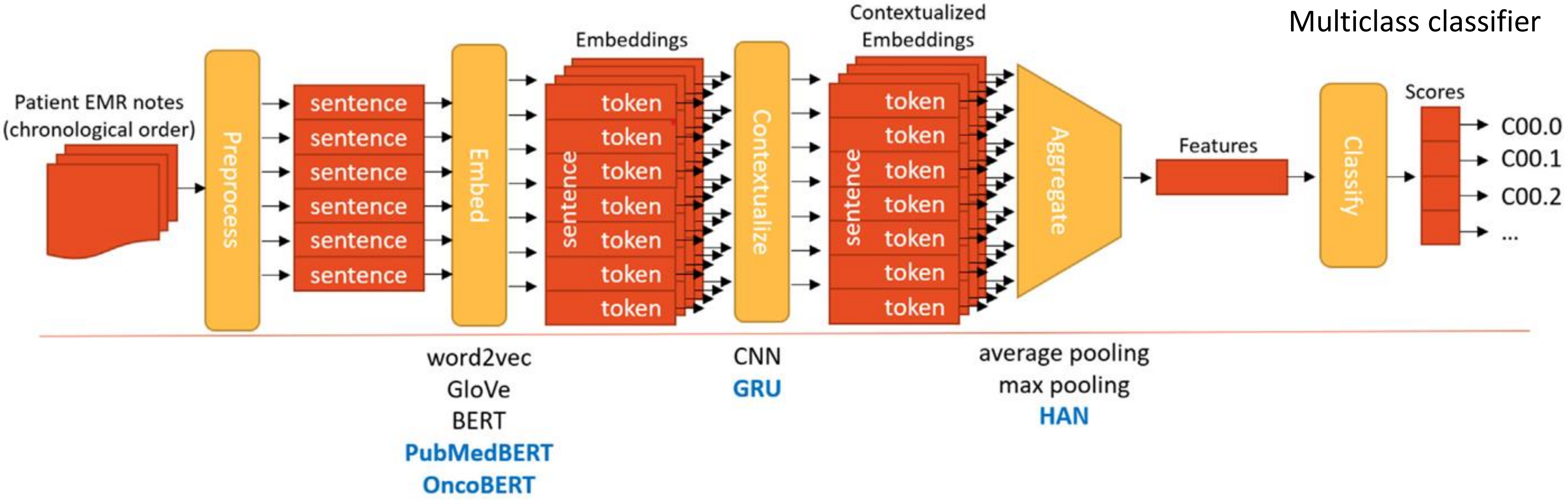


Figure 6. A general neural architecture for medical abstraction

Clinical documents are concatenated by chronological order and converted into a token sequence, which is then transformed into a sequence of neural vectors by the embedding and contextualization modules, before being converted into a fixed-length feature vector by an aggregation module for final classification.

- Note

  - OncoGlove:

    100-dimensional GloVe embedding pretrained on oncology notes.

  - PubMedBERT:

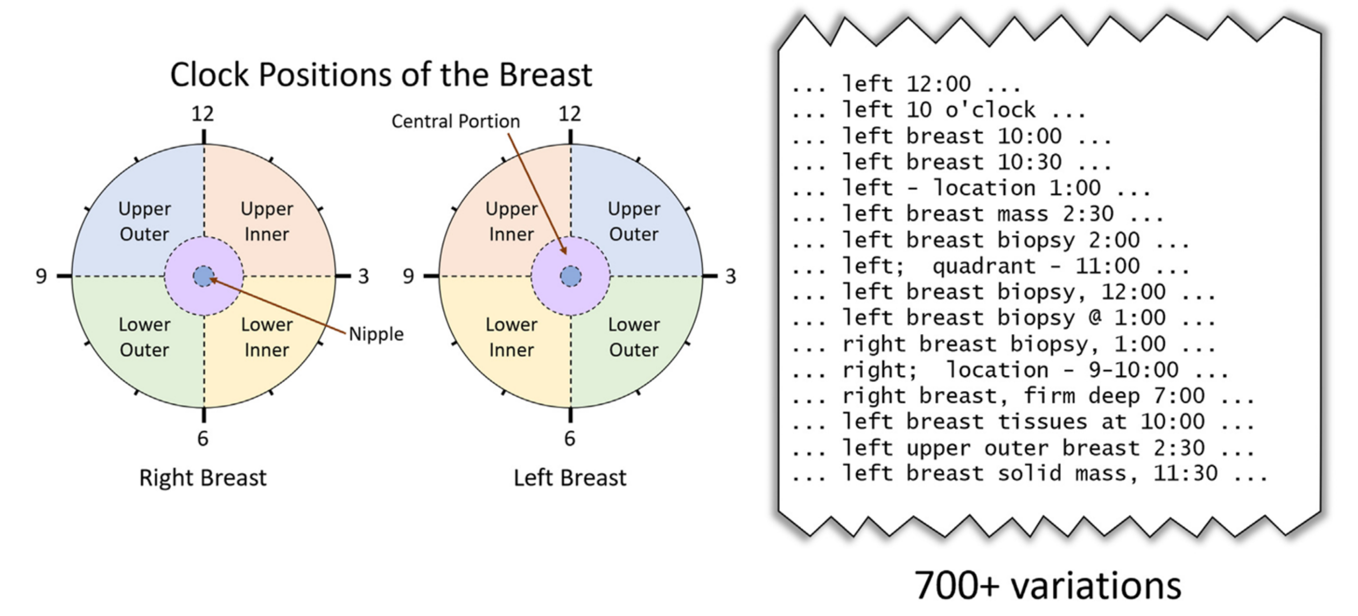
    Pretrained from scratch using abstracts from PubMed and full-text articles from PubMedCentral .

  - OncoBERT :

    Fine-tuned BERT with unstructured EHR from breast cancer, prostate cancer and glioma patients (over 1 million patients).

- Challenge in medical abstraction with NLP

  -myriad variations



```
... left 12:00 ...
... left 10 o'clock ...
... left breast 10:00 ...
... left breast 10:30 ...
... left - location 1:00 ...
... left breast mass 2:30 ...
... left breast biopsy 2:00 ...
... left;  quadrant - 11:00 ...
... left breast biopsy, 12:00 ...
... left breast biopsy @ 1:00 ...
... right breast biopsy, 1:00 ...
... right;  location - 9-10:00 ...
... right breast, firm deep 7:00 ...
... left breast tissues at 10:00 ...
... left upper outer breast 2:30 ...
... left breast solid mass, 11:30 ...
```

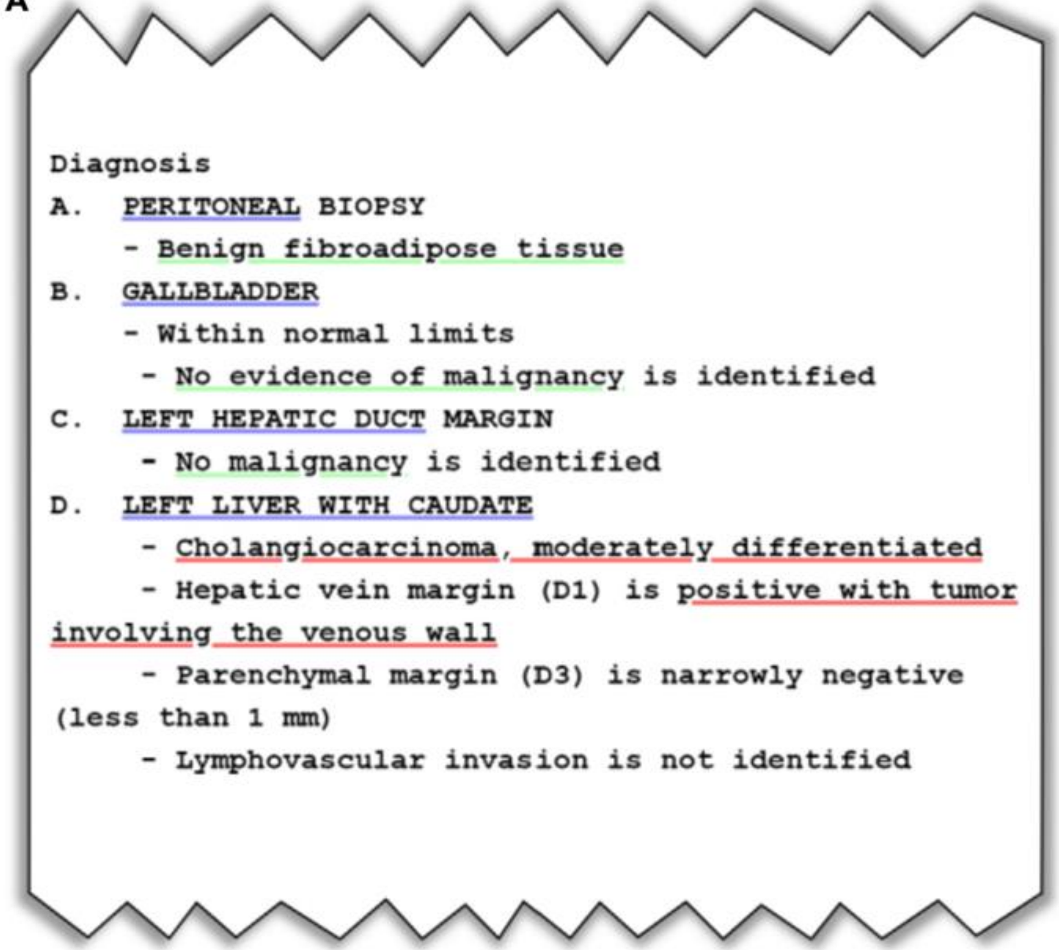
700+ variations

Figure 5. Variations in medical abstraction

Relevant information for medical abstraction may manifest in myriad variations, as seen in specification of tumor site in breast cancer with laterality and clockwise position.

- name entity recognition is not enough
  : many candidate sites may be present

A



Diagnosis

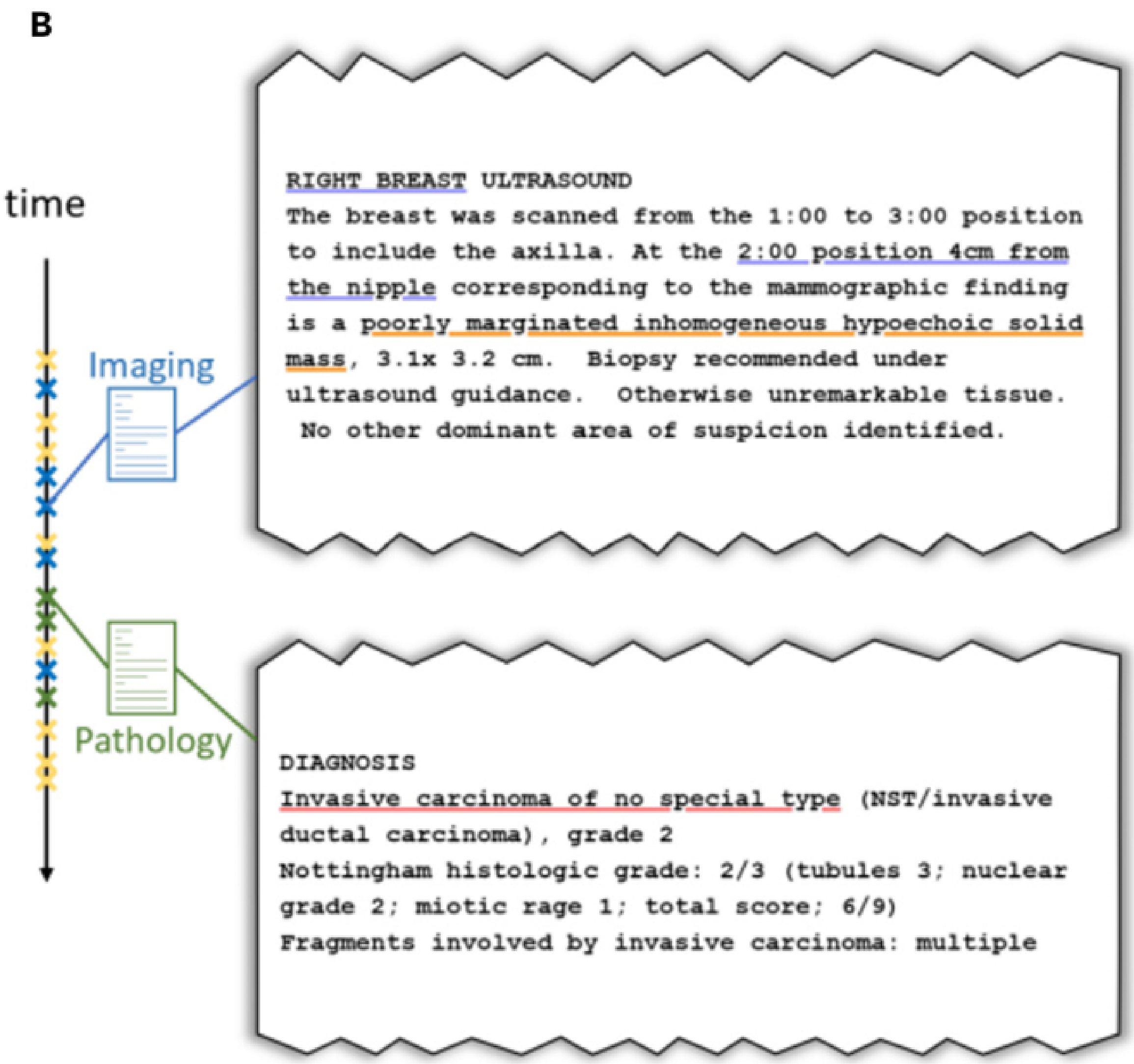
A. PERITONEAL BIOPSY
- Benign fibroadipose tissue

B. GALLBLADDER
- Within normal limits
- No evidence of malignancy is identified

C. LEFT HEPATIC DUCT MARGIN
- No malignancy is identified

D. LEFT LIVER WITH CAUDATE
- Cholangiocarcinoma, moderately differentiated
- Hepatic vein margin (D1) is positive with tumor involving the venous wall
- Parenchymal margin (D3) is narrowly negative (less than 1 mm)
- Lymphovascular invasion is not identified

- Abstraction may require information integration across multiple clinical documents, different times



RIGHT BREAST ULTRASOUND
The breast was scanned from the 1:00 to 3:00 position to include the axilla. At the 2:00 position 4cm from the nipple corresponding to the mammographic finding is a poorly marginated inhomogeneous hypoechoic solid mass, 3.1x 3.2 cm. Biopsy recommended under ultrasound guidance. Otherwise unremarkable tissue. No other dominant area of suspicion identified.

DIAGNOSIS
Invasive carcinoma of no special type (NST/invasive ductal carcinoma), grade 2
Nottingham histologic grade: 2/3 (tubules 3; nuclear grade 2; miotic rage 1; total score; 6/9)
Fragments involved by invasive carcinoma: multiple

The location is described in an imaging report, whereas the positive diagnosis is documented in a pathology report.

- Abstraction may require information integration across multiple clinical documents.

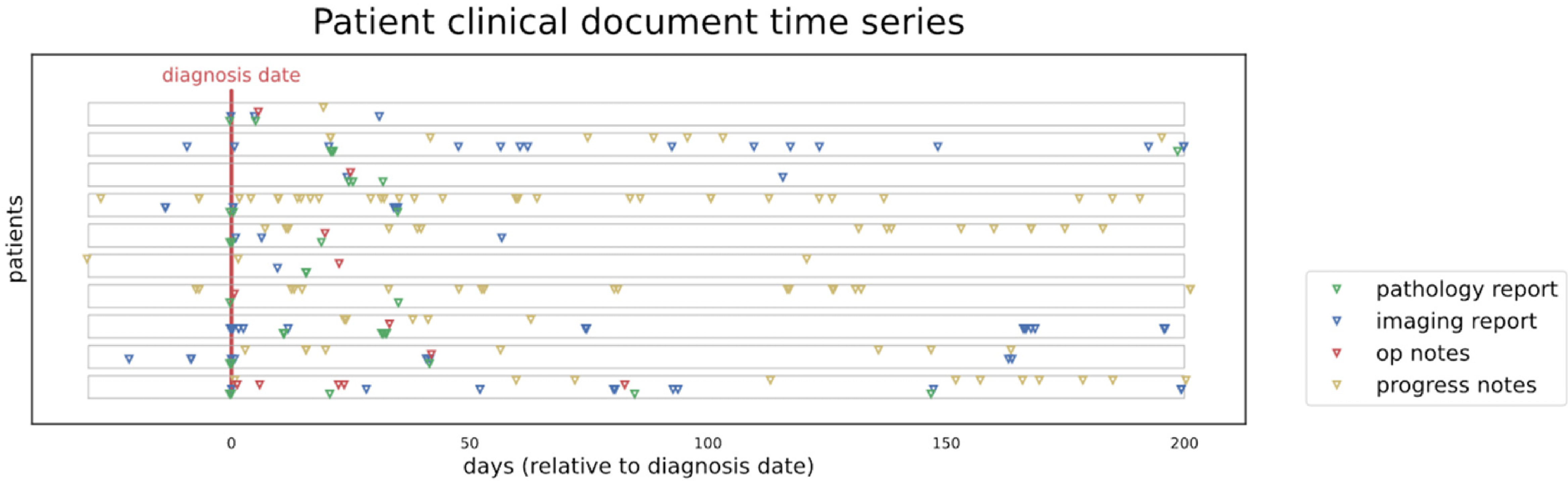


Figure 4. Patient clinical document time series

Patients with cancer typically have many clinical documents for a tumor diagnosis, with key information scattered among these documents.

Results

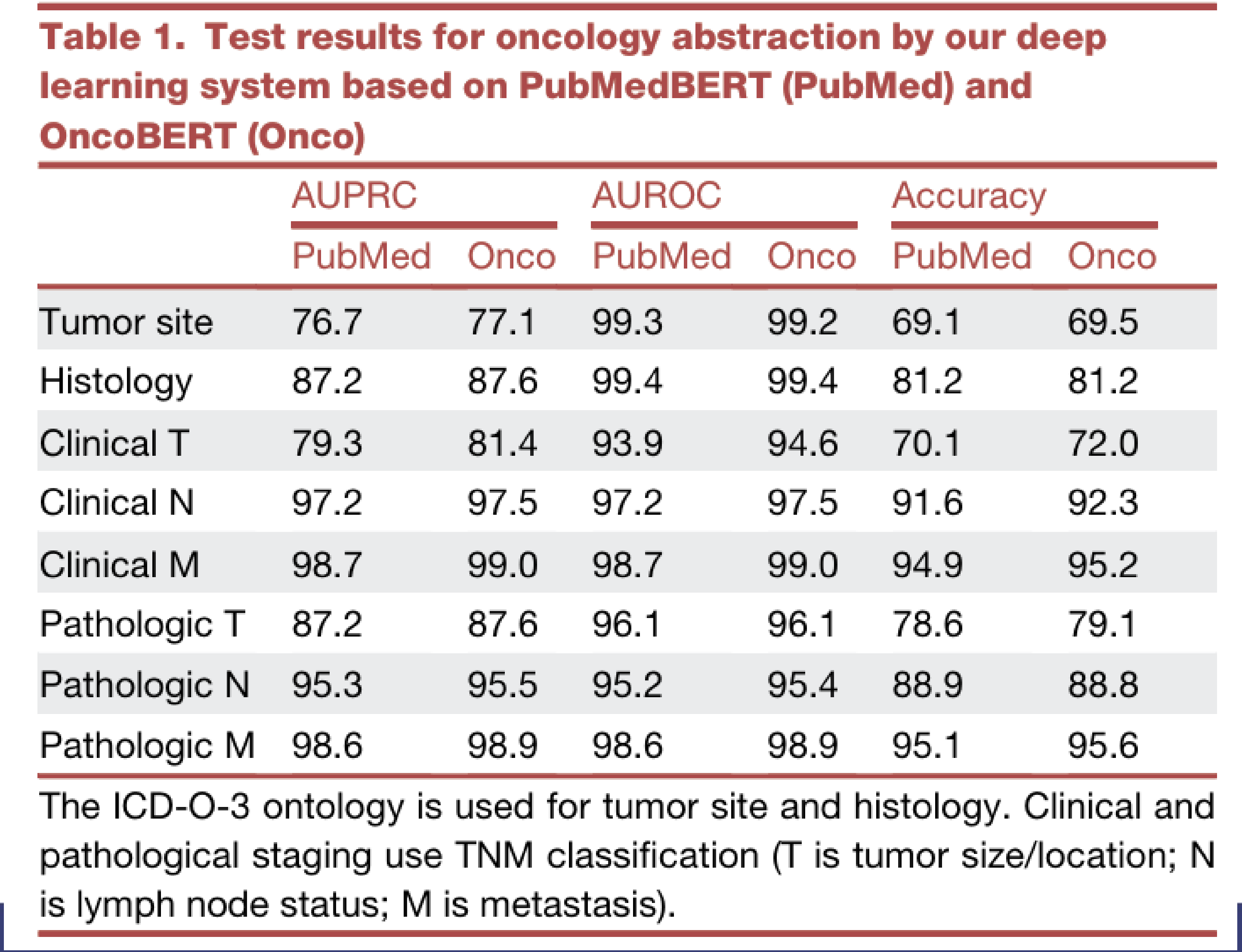


Table 1. Test results for oncology abstraction by our deep learning system based on PubMedBERT (PubMed) and OncoBERT (Onco)

	AUPRC		AUROC		Accuracy	
	PubMed	Onco	PubMed	Onco	PubMed	Onco
Tumor site	76.7	77.1	99.3	99.2	69.1	69.5
Histology	87.2	87.6	99.4	99.4	81.2	81.2
Clinical T	79.3	81.4	93.9	94.6	70.1	72.0
Clinical N	97.2	97.5	97.2	97.5	91.6	92.3
Clinical M	98.7	99.0	98.7	99.0	94.9	95.2
Pathologic T	87.2	87.6	96.1	96.1	78.6	79.1
Pathologic N	95.3	95.5	95.2	95.4	88.9	88.8
Pathologic M	98.6	98.9	98.6	98.9	95.1	95.6

The ICD-O-3 ontology is used for tumor site and histology. Clinical and pathological staging use TNM classification (T is tumor size/location; N is lymph node status; M is metastasis).

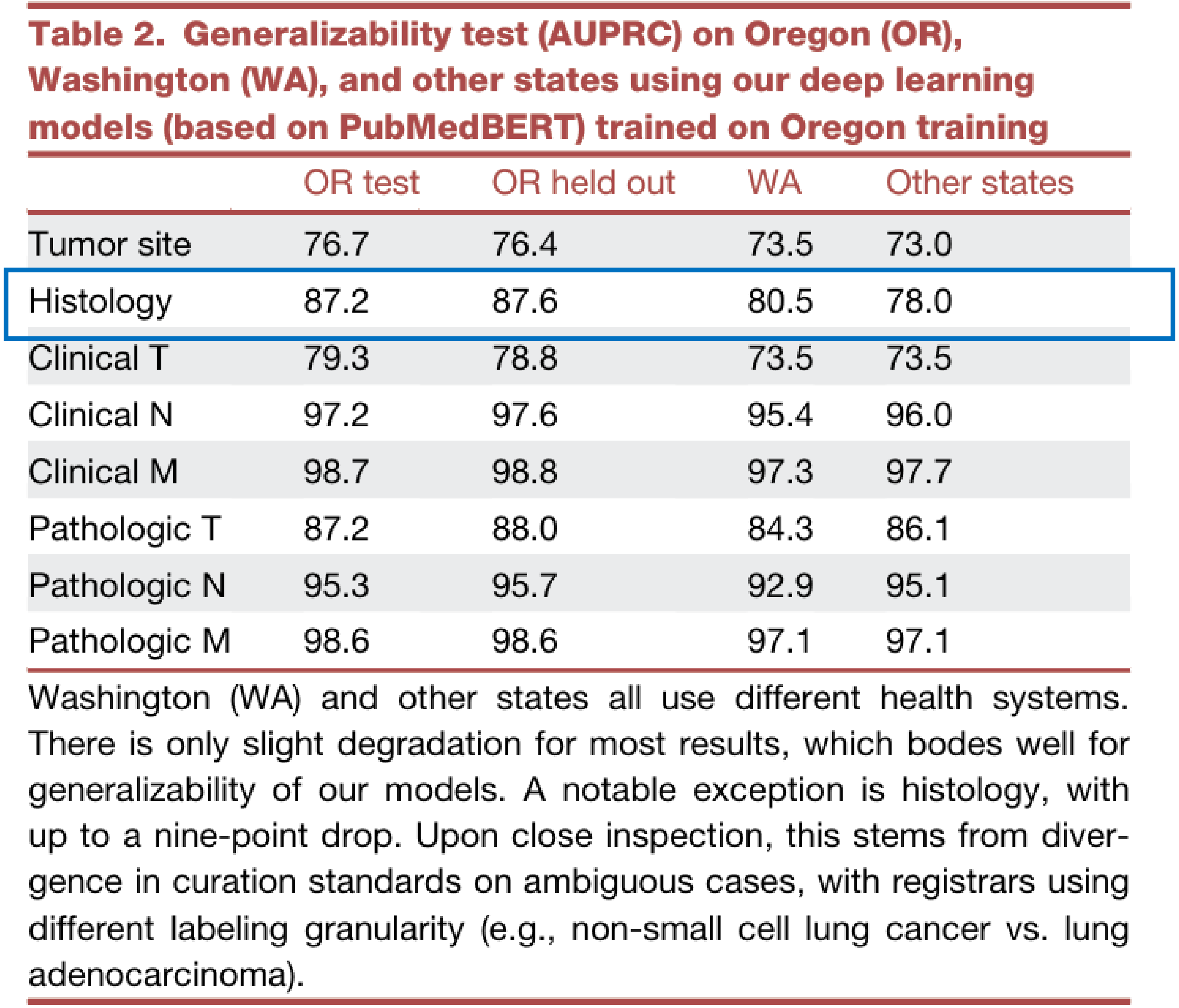


Table 2. Generalizability test (AUPRC) on Oregon (OR), Washington (WA), and other states using our deep learning models (based on PubMedBERT) trained on Oregon training

	OR test	OR held out	WA	Other states
Tumor site	76.7	76.4	73.5	73.0
Histology	87.2	87.6	80.5	78.0
Clinical T	79.3	78.8	73.5	73.5
Clinical N	97.2	97.6	95.4	96.0
Clinical M	98.7	98.8	97.3	97.7
Pathologic T	87.2	88.0	84.3	86.1
Pathologic N	95.3	95.7	92.9	95.1
Pathologic M	98.6	98.6	97.1	97.1

Washington (WA) and other states all use different health systems. There is only slight degradation for most results, which bodes well for generalizability of our models. A notable exception is histology, with up to a nine-point drop. Upon close inspection, this stems from divergence in curation standards on ambiguous cases, with registrars using different labeling granularity (e.g., non-small cell lung cancer vs. lung adenocarcinoma).

different labeling granularity

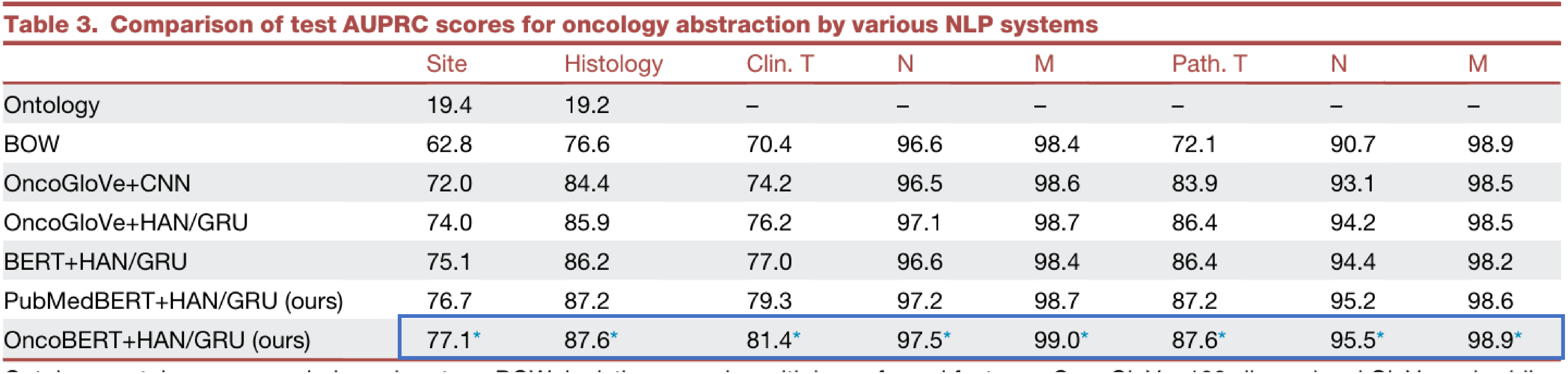


Table 3. Comparison of test AUPRC scores for oncology abstraction by various NLP systems

	Site	Histology	Clin. T	N	M	Path. T	N	M
Ontology	19.4	19.2	–	–	–	–	–	–
BOW	62.8	76.6	70.4	96.6	98.4	72.1	90.7	98.9
OncoGloVe+CNN	72.0	84.4	74.2	96.5	98.6	83.9	93.1	98.5
OncoGloVe+HAN/GRU	74.0	85.9	76.2	97.1	98.7	86.4	94.2	98.5
BERT+HAN/GRU	75.1	86.2	77.0	96.6	98.4	86.4	94.4	98.2
PubMedBERT+HAN/GRU (ours)	76.7	87.2	79.3	97.2	98.7	87.2	95.2	98.6
OncoBERT+HAN/GRU (ours)	77.1*	87.6*	81.4*	97.5*	99.0*	87.6*	95.5*	98.9*

Ablation study

- Adding radiology reports on top of pathology reports increased the AUPRC by 3.4 absolute points for tumor site extraction.
- Adding the operative notes providing an additional 1-point gain.
- For pathological staging, however, a larger window is helpful, using[-30,90] days as input improves the AUPRC by 4 absolute points for pathological T staging.

Case findings

- Cancer providers are obligated to submit abstraction for these patients to the registry within a time limit.
- 62,090 positive and 8,460 negative (non cancer) patients.
- train/development/test : 60% / 20% / 20%
- For patients with cancer, the case finding decision is deemed correct if the first day of positive classification is within[-7,30] days of diagnosis.

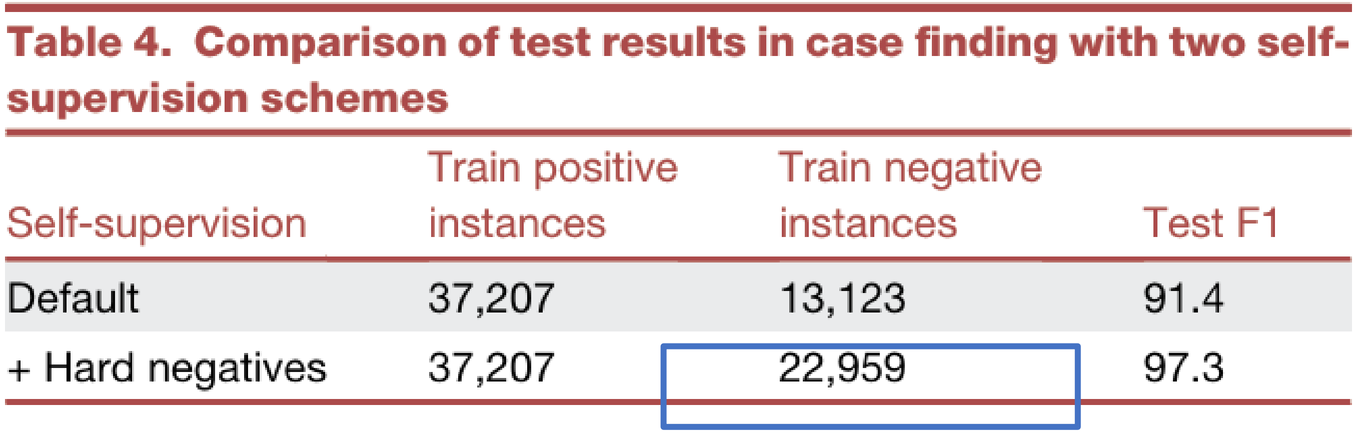


Table 4. Comparison of test results in case finding with two self-supervision schemes

Self-supervision	Train positive instances	Train negative instances	Test F1
Default	37,207	13,123	91.4
+ Hard negatives	37,207	22,959	97.3

Add negative instances comprise of randomly chosen

- days among non-cancer patients.
- Days at least a week before diagnosis (up to a year before) among patients with cancer

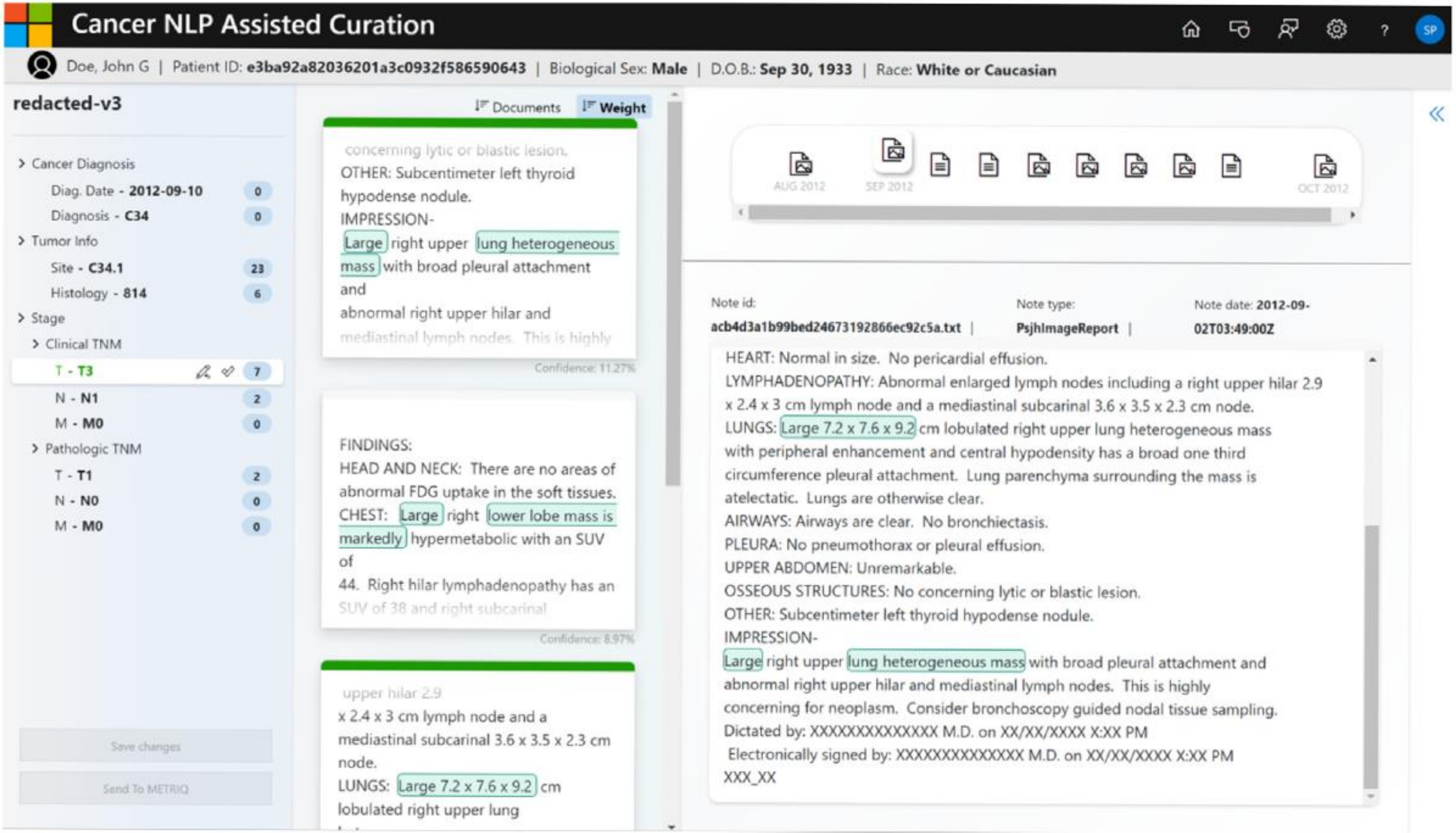
Discussions

Error analysis

- Manual analysis on sample errors.
  - annotation inconsistency
  - missing notes.
- By analyzing 50 error examples for tumor site classification, we found that a significant proportion of incorrect annotations
  : real test AUPRC is about 91.6 (vs.76.7).

Assisted curation

- The attention mechanism in transformer-based models provides a straightforward approach to identify extraction rationale.
- However, there is no guarantee that attention provides explanation.
- A research prototype that we have developed for assisted curation, which is in test use by selected clinical users.
- In preliminary studies, tumor registrars can verify a candidate extraction in 1–2 min, either ascertaining its correctness or fixing the label in the interface.



Cancer NLP Assisted Curation

Doe, John G | Patient ID: **e3ba92a82036201a3c0932f586590643** | Biological Sex: **Male** | D.O.B.: **Sep 30, 1933** | Race: **White or Caucasian**

redacted-v3

Cancer Diagnosis	
Diag. Date - **2012-09-10**	0
Diagnosis - **C34**	0
Tumor Info	
Site - **C34.1**	23
Histology - **814**	6
Stage	
Clinical TNM	
T - **T3**	7
N - **N1**	2
M - **M0**	0
Pathologic TNM	
T - **T1**	2
N - **N0**	0
M - **M0**	0

Save changes

Send To METRIQ

Documents | **Weight**

concerning lytic or blastic lesion.
OTHER: Subcentimeter left thyroid hypodense nodule.
IMPRESSION-
Large right upper lung heterogeneous mass with broad pleural attachment and
abnormal right upper hilar and mediastinal lymph nodes. This is highly

Confidence: 11.27%

FINDINGS:
HEAD AND NECK: There are no areas of abnormal FDG uptake in the soft tissues.
CHEST: Large right lower lobe mass is markedly hypermetabolic with an SUV of
44. Right hilar lymphadenopathy has an SUV of 38 and right subcarinal

Confidence: 8.97%

upper hilar 2.9
x 2.4 x 3 cm lymph node and a mediastinal subcarinal 3.6 x 3.5 x 2.3 cm node.
LUNGS: Large 7.2 x 7.6 x 9.2 cm lobulated right upper lung

AUG 2012 SEP 2012 OCT 2012

Note id: **acb4d3a1b99bed24673192866ec92c5a.txt** | Note type: **PsjhImageReport** | Note date: **2012-09-02T03:49:00Z**

HEART: Normal in size. No pericardial effusion.
LYMPHADENOPATHY: Abnormal enlarged lymph nodes including a right upper hilar 2.9 x 2.4 x 3 cm lymph node and a mediastinal subcarinal 3.6 x 3.5 x 2.3 cm node.
LUNGS: Large 7.2 x 7.6 x 9.2 cm lobulated right upper lung heterogeneous mass with peripheral enhancement and central hypodensity has a broad one third circumference pleural attachment. Lung parenchyma surrounding the mass is atelectatic. Lungs are otherwise clear.
AIRWAYS: Airways are clear. No bronchiectasis.
PLEURA: No pneumothorax or pleural effusion.
UPPER ABDOMEN: Unremarkable.
OSSEOUS STRUCTURES: No concerning lytic or blastic lesion.
OTHER: Subcentimeter left thyroid hypodense nodule.
IMPRESSION-
Large right upper lung heterogeneous mass with broad pleural attachment and abnormal right upper hilar and mediastinal lymph nodes. This is highly concerning for neoplasm. Consider bronchoscopy guided nodal tissue sampling.
Dictated by: XXXXXXXXXXXXX M.D. on XX/XX/XXXX X:XX PM
Electronically signed by: XXXXXXXXXXXXX M.D. on XX/XX/XXXX X:XX PM
XXX_XX

Figure 3. Cancer NLP-assisted curation system

Our cancer-assisted curation system. Left: extracted oncology attributes. Middle: extraction rationale based on attention weights. Right: full notes. Patient information has been deidentified.

Fairness

- We conducted a performance evaluation for each gender and ethnicity subgroup in the test set.
- Equal performance was observed on almost all scenarios, except for tumor site abstraction on the subgroup ''Native Hawaiian or Other Pacific Islander.''
- Investigate : ethnicity information in the notes, ethnic stereotypes and biases may be reflected in pretrained embeddings.
- We found that less than 2% of top-attention tokens were ethnicity-related tokens.
- The most likely explanation is random fluctuation stemming from the very small sample size (29 patients).

Conclusion

- Manual curation of complex clinical records and EMR data is expensive and time consuming.
- By applying our NLP system to all patients, we instantly expand structured RWD for the network by an order of magnitude.
- In future work, we plan to expand the scope of curation by applying self-supervised learning to extracting other key information for real-world evidence, such as treatments and key clinical outcomes such as response to immunotherapy.

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

THANK YOU

Wisdom of the Land