Explainable automated medical coding and weakly supervised rare disease identification from clinical notes

HANG DONG, RESEARCH FELLOW, USHER INSTITUTE
UNIVERSITY OF EDINBURGH
HANG.DONG@ED.AC.UK

Overview

- Use case 1: Automated medical coding
  - Method – explainability with attention mechanisms
  - Results
  - Demonstration

- Use case 2: Rare disease cohort selection
  - Method – improve SemEHR with rule-based weak supervision
  - Preliminary results

Use case 1: Automated Medical Coding

Assigning medical codes (e.g. ICD codes) for clinical notes, to facilitate the work of coding staff.

In NHS Lothian, each staff codes 60 discharge summaries per day (approx. 7min/doc), and the whole team codes 20 000 every month (Interview by Rannikmäe, 2016).

Challenging task (Baumel *et al.*, 2018):

- **Large number of labels:** about 9000 ICD 9 code used in MIMIC-III (US discharge summaries in ICU)
- **Multi-label setting**: 16 code per discharge summary on average in MIMIC-III
- **Long documents**: 2000 tokens in a discharge summary on average in MIMIC-III

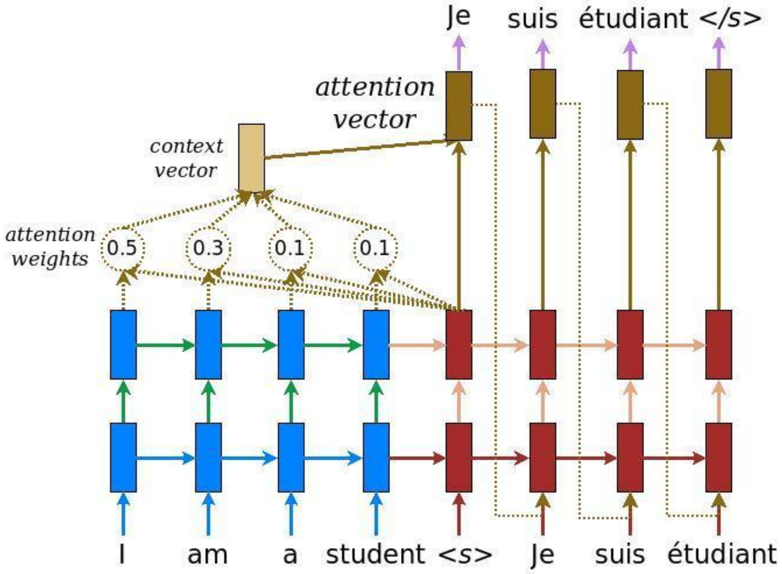
Explainable ML for EHR with attention mechanisms

Explainability is "the ability to explain what happened when the model made a decision, in terms that a person understands" (Geis *et al.*, Ethics of Artificial Intelligence in Radiology, 2019, p. 438)

Methods for Explainable ML with EHR data: feature interaction & importance, **attention mechanisms**, data dimensionality reduction, knowledge distillation and rule extraction, intrinsically interpretable models (Payrovnaziri *et al.*, JAMIA, 2020).

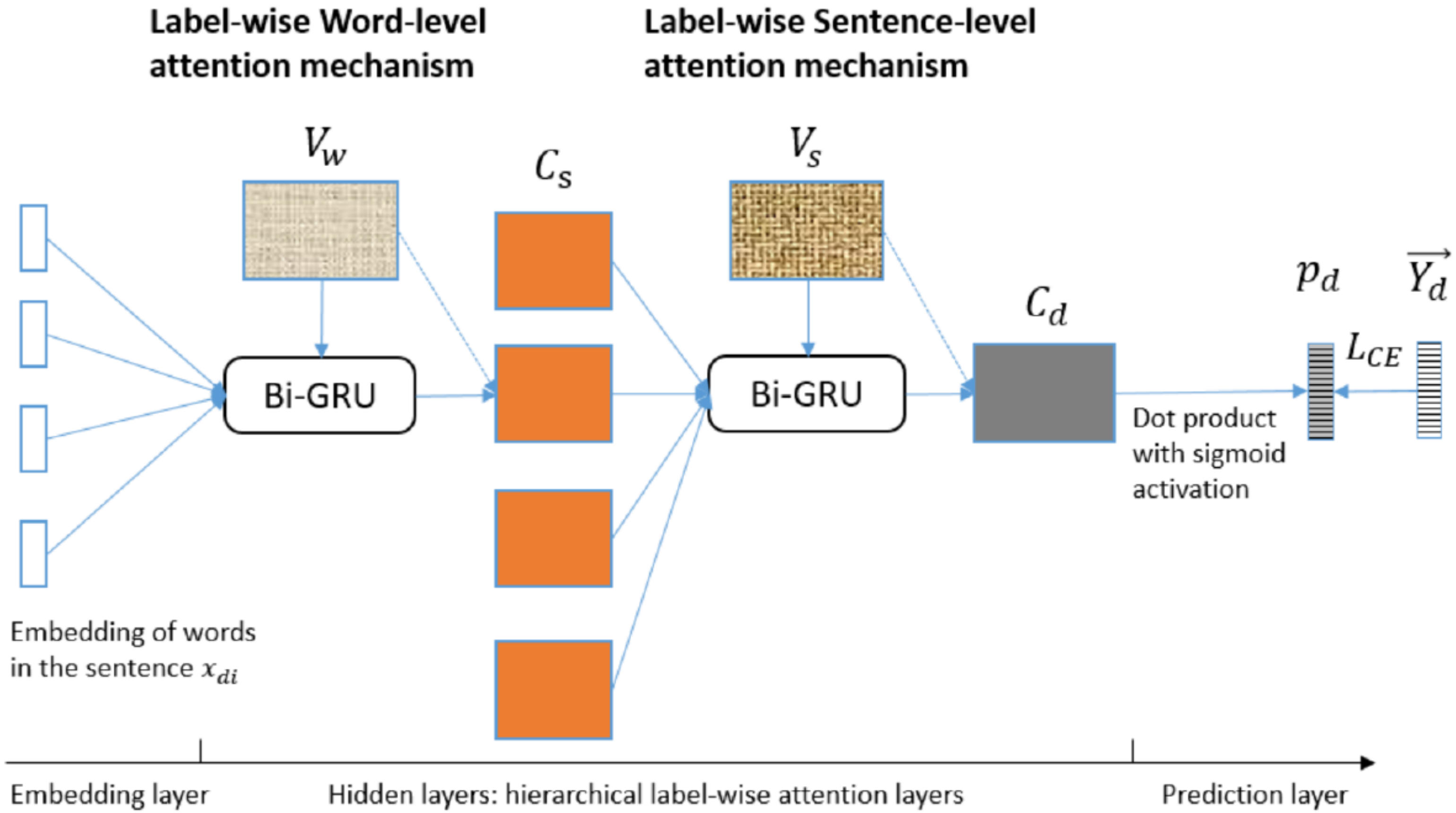
Attention mechanism (Bahdanau *et al.*, 2014) is a major approach considered to enhance the explainability of *deep learning* models with EHR data.

Attention mechanism usually learns a peaky distribution of the input sequence regarding the prediction task.



Acknowledgement to the Image from (Synced, 2017) https://medium.com/syncedreview/a-brief-overview-of-attention-mechanism-13c578ba9129

Hierarchical Label-wise Attention Network (HLAN) with Label Embedding (LE) Initialisation



We further inject label embedding to the attention layers and the prediction layer.

Image from Dong H, Suárez-Paniagua V, Whiteley W, Wu H. Explainable Automated Coding of Clinical Notes using Hierarchical Label-wise Attention Networks and Label Embedding Initialisation. Journal of Biomedical Informatics. 2021 Mar 9;103728. Preprint on arXiv: https://arxiv.org/abs/2010.15728

Datasets

MIMIC-III and MIMIC-III-50 (top-50 code prediction), Medical Information Mart for Intensive Care, preprocessed as in (Mullenbach *et al.*, 2018).

Discharge summaries, in average 2000 tokens per document, all padded to 2500 tokens.

MIMIC-III-shielding, discharge summaries with ICD-9 code corresponding to the NHS ICD-10 code to select high-risk patients for shielding during COVID-19.

Numbers	MIMIC-III	MIMIC-III-50	MIMIC-III-shielding
Training documents	47,724	8,067	4,574
Labels	**8,922**	50	20
Labels per document	15.88	5.69	1.08

MIMIC-III Dataset from: Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Liwei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. 2016. MIMIC-III, a freely accessible critical care database. Scientific Data, 3(1):1–9.

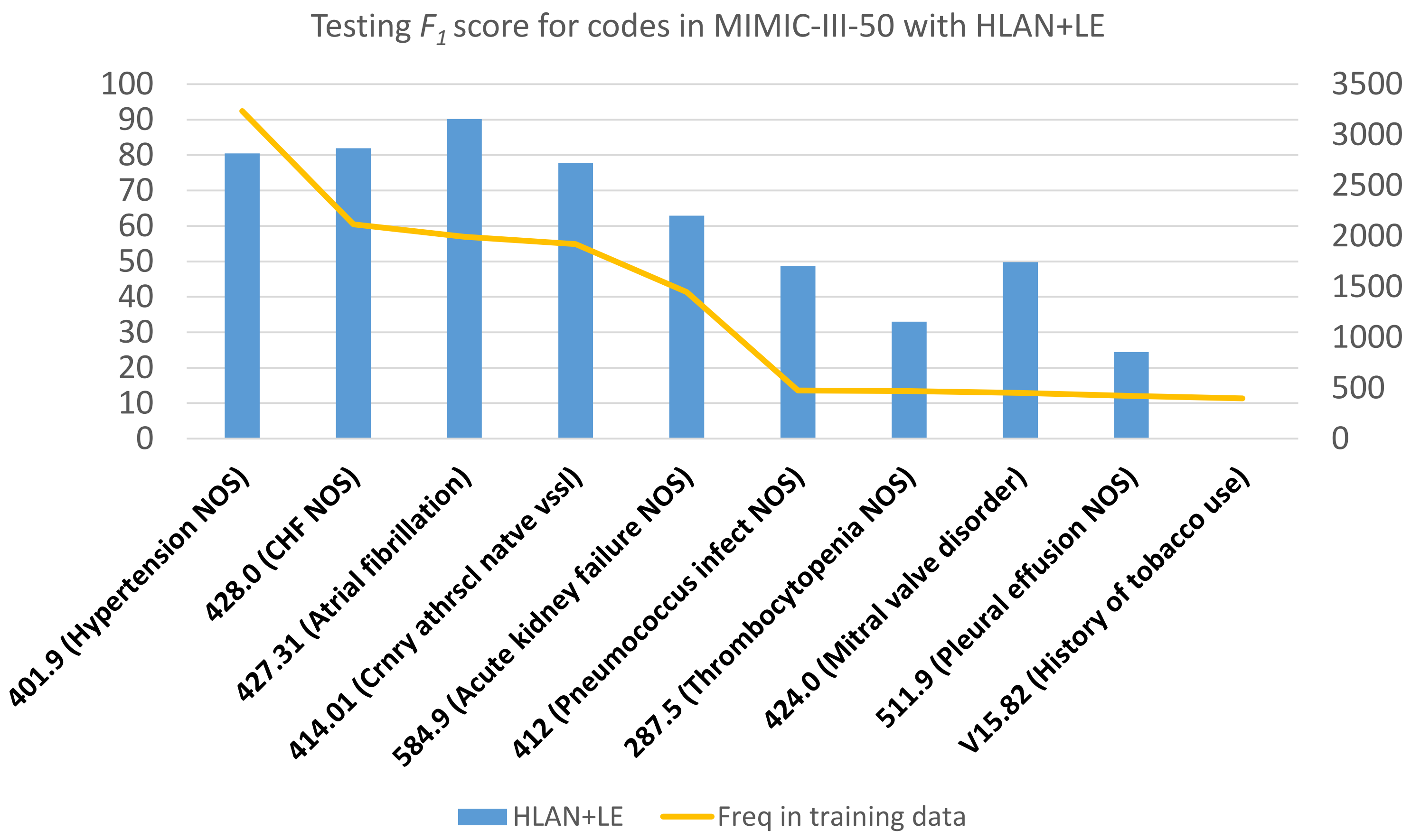
Results and Demonstration

Testing results averaged 10 times with random seeds

Models (MIMIC-III-50)	Micro-level precision	Micro-level recall	Micro-level F1
CNN	55.6%	***71.2%***	62.4%
CNN+att	70.9%	53.1%	60.7%
Bi-GRU	58.1%	45.8%	51.2%
HAN	68.2%	52.9%	59.4%
HA-GRU	69.5%	48.7%	57.2%
HLAN+LE (ours)	***73.2%***	56.9%	***64.0%***

Italics: p<0.05 to the second best result

Time for demonstration!



A visualization of explainable coding with HLAN

For 427.31	For 428.0	Document #24 in MIMIC-III-50										
0.02	0.01	admission	date	discharge	date	service	surgery					
0.01	0.01	allergies	patient	recorded	as	having	no	known	allergies	to	drugs	attending
0.01	0.02	major	surgical	or	invasive	procedure	ex	lap	r	hemicolectomy	mucous	fistula
0	0	history	of	present	illness	age	over	f	presented	to	location	un
0	0.01	admitted	to	hospital1	and	taken	directly	to	or	upon	arrival	
0.41	0.8	past	medical	history	pmhx	a	fib	aortic	stenosis	chf	last	ef
0	0	pertinent	results	00pm	blood	wbc	rbc	hgb	hct	mcv	mch	mchc
0	0.01	brief	hospital	course	age	over	f	transferred	from	location	un	and
0	0.02	patient	was	taken	directly	to	the	operating	room	for	an	exploratory
0	0	she	underwent	a	right	hemicolectomy	mucous	fistula	ileostomy	gj	tube	placement
0.01	0.01	fluid	balance	intraoperatively	included	units	ffp	units	plt	u	nits	prbc
0	0	patient	was	kept	intubated	and	taken	directly	to	the	surgical	intensive
0	0.02	she	required	maximum	pressor	support	to	maintain	sufficient	cardiac	index	
0	0.03	patient	did	show	signs	of	distal	ischemia	to	extremities	by	the
0.01	0	family	meeting	at	latter	evening	decided	to	make	patient	cmo	patient
0.01	0	medications	on	admission	last	name	un	amlodipine	mg	qd	benarepril	mg
0.49	0	discharge	diagnosis	cardiopulmonary	arrest	perforated	colon	atrial	fibrillation	ventilatory	support	discharge

Sentence-level attention weights

427.31 (Atrial Fibrillation)

428.0 (Congestive heart failure, unspecified)

Use Case 2: Rare Disease Cohort Selection

Rare diseases are likely to be missed in ICD coding.

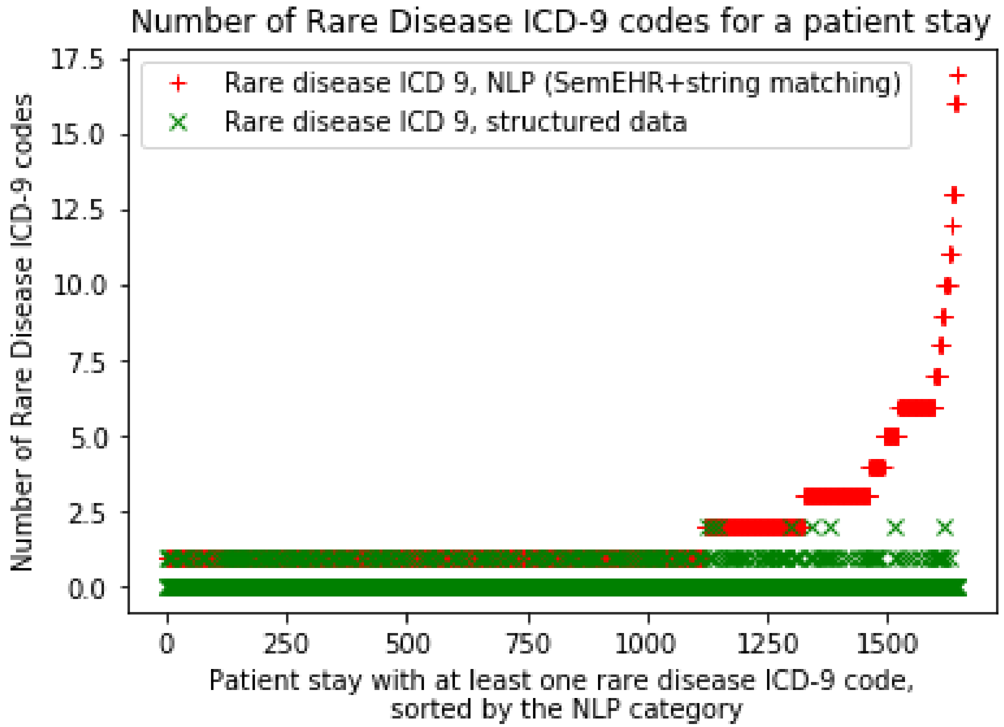
- Only 500 rare diseases have a specific ICD 10 code (Bearryman, 2016).
  - *cf.* 3,000 rare disease in SNOMED CT and 7,000 in Orpha.net
- Only 6 diagnosis code for each admission episode, with **priority** given to common diseases and limited **time** for coding.

NLP can surface rare diseases for cohort selection.

Ontology-based approach:

Free-text -> UMLS -> ORDO (-> ICD)

Preliminary Results with 10,000 MIMIC-III discharge summaries using SemEHR and string matching



NER+L tool: SemEHR (from text to UMLS)

Bio-YODIE (Gorrell *et al.*, 2018) is the main NLP Pipeline in SemEHR.

Bio-YODIE is fast, using a gazetteer-based approach to match texts to gazetters (e.g. UMLS terms and synonyms).

Limitation: no contextual disambiguation and especially fail in matching abbreviations.

For example, some False Positives in SemEHR:

(i) His temporary HD line was pulled.
(ii) … male with ESRD on HD …
(iii) Discharge Medications … 3. Asacol HD 800 mg Tablet …
(iv) CT scan on HD9 showed …

All "HD"s were **falsely** identified as **Huntingdon Disease (has synonym HD).**

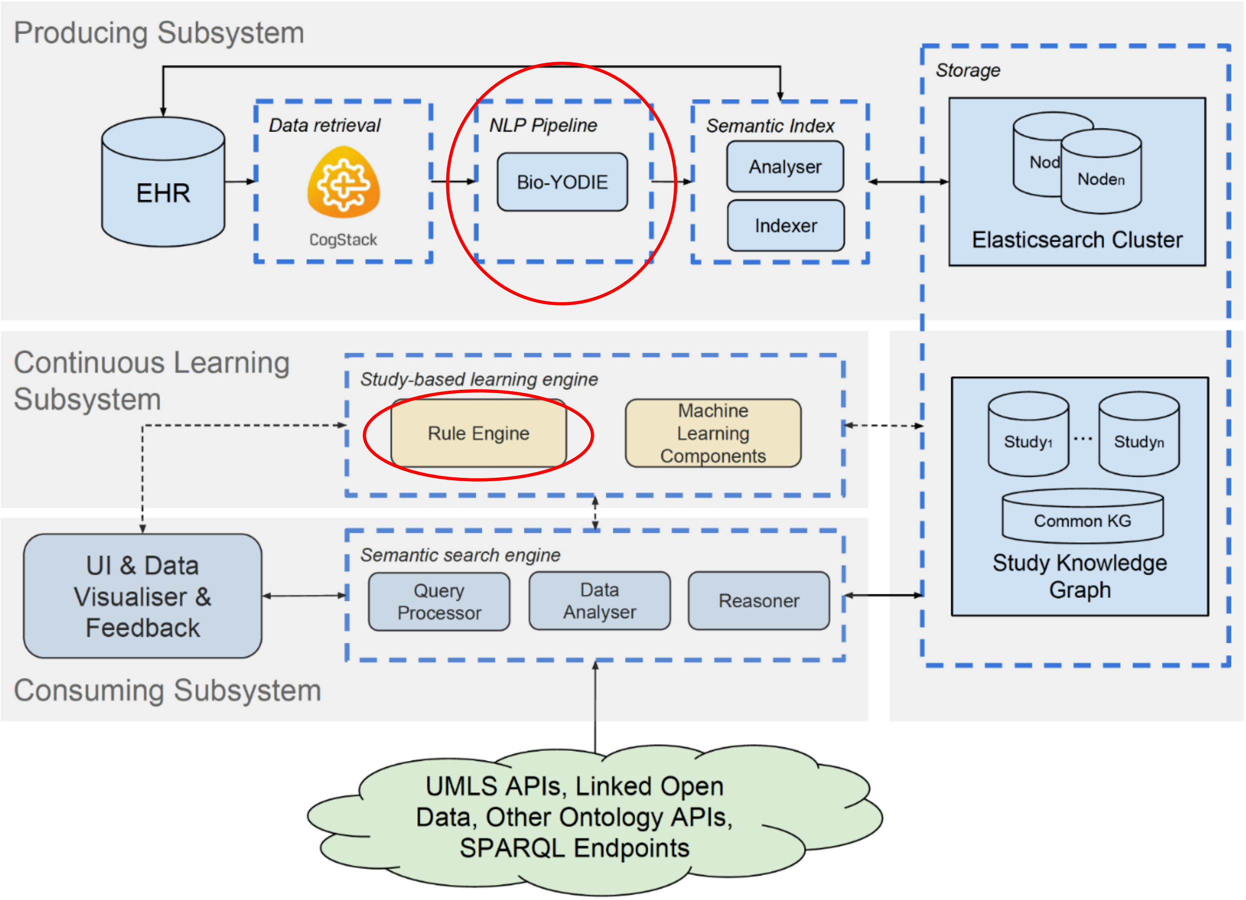


Image from Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research*. Journal of the American Medical Informatics Association. 2018 May 1;25(5):530–7.

Weak supervision

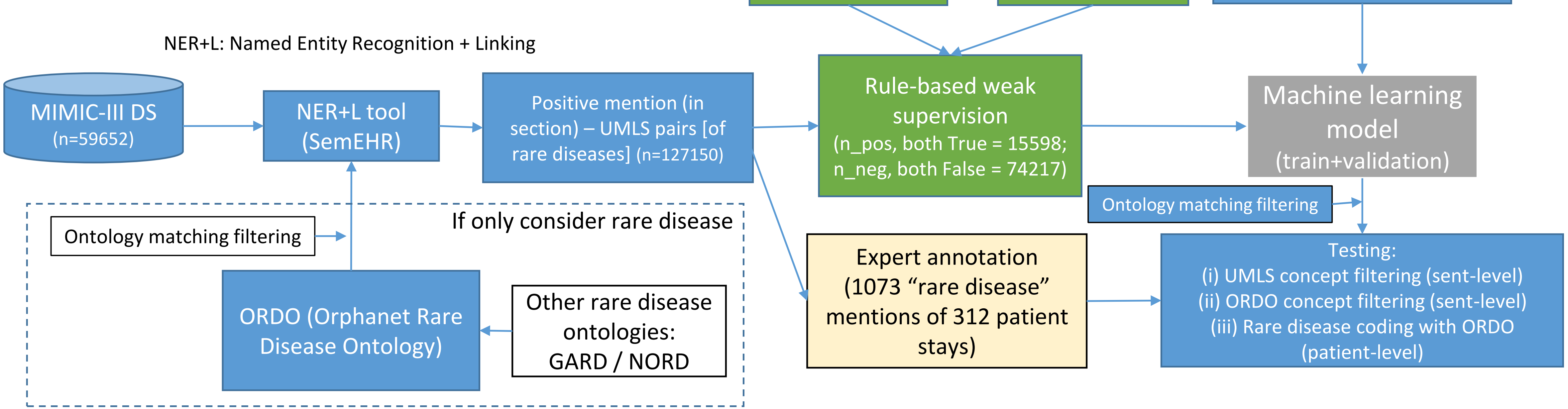
Weak supervision: get low-quality labels efficiently (using noisy, limited, or imprecise sources) without hand-labelling from subject matter experts (Ratner *et al.*, 2019).

Sources for weak supervision (Ratner *et al.*, 2019):

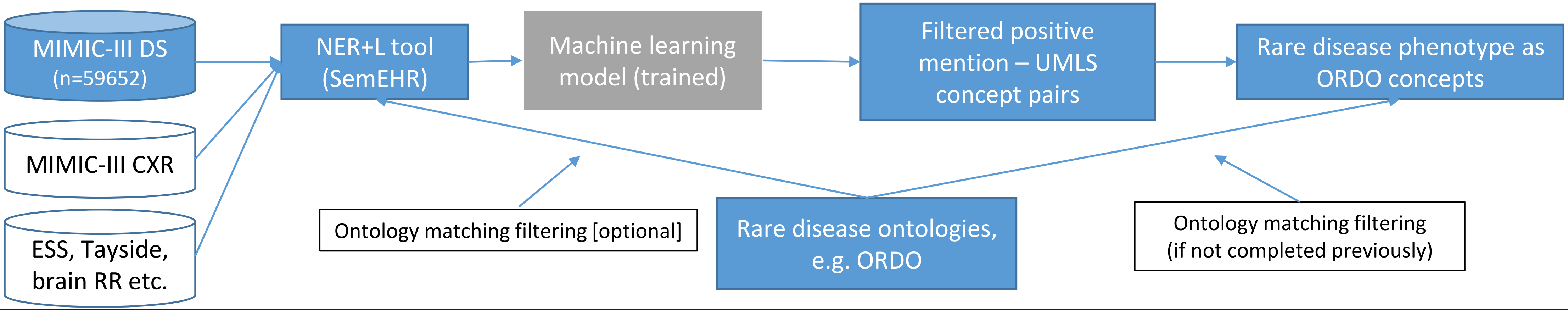
- heuristics, rules
- crowdsourcing
- Knowledge Bases – distant supervision
- existing resources, e.g. automated tools, pre-trained models
- etc.

We use SemEHR with rules to generate weak labels.

Training stage:
whether a mention-UMLS pair indicates a rare phenotype



Inference stage: filtering with the model



Results on rare disease UMLS filtering

Results based on MIMIC-III discharge summaries, randomly sampled 200 mention-to-UMLS-disease pairs identified by SemEHR (see examples below).

UMLSs were filtered by those matched to Orphanet Rare Disease Ontology (ORDO).

Rare disease UMLS filtering	precision	recall	F1
Google Healthcare Natural Language API (released in Nov 2020)	76.9%	84.5%	80.5%
SemEHR	35.5%	**100.0%** (reference)	52.4%
SemEHR + two rules ("OR" operation)	85.7%	93.0%	89.2%
SemEHR + rule-based weak supervision with BlueBERT (ours)	**91.7%**	93.0%	**92.3%**

Text: History of pneumonia in [**2097**] complicated by ARDS. [**2099**] tetanus and pneumovax vaccination with a negative PPD at that time." –

False

UMLS: C0039614 Tetanus

Text: Past Medical History:
7. Heparin induced thrombocytopenia in [**9-/2128**]

True

UMLS: C0272285 HIT

Conclusions and Future studies

- **Explainable** Automated medical coding through deep learning and attention mechanisms.
- Rare disease identification with ontologies, SemEHR, and rule-based **weak** supervision.

Areas for future studies:

- Generalise the models for data from the UK
  - Edinburgh Stroke Study + NHS Tayside brain imaging reports (already acquired)
  - DataLoch (Edinbugh and South East Scotland region)
- Identify issues and gaps to deploy the coding system in the NHS:
  - Connect to NHS staff, researchers, clinical software companies. (Please contact me if interested)

Preprint (accepted to JBI): https://arxiv.org/abs/2010.15728
GitHub: https://github.com/acadTags/Explainable-Automated-Medical-Coding

References

- Baumel T, Nassour-Kassis J, Cohen R, Elhadad M, Elhadad N. Multi-label classification of patient notes: case study on ICD code assignment. In: Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence. 2018. arXiv:1709.09587. https://arxiv.org/pdf/1709.09587.pdf
- Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. Radiology. 2019 Oct 1;293(2):436–40. https://doi.org/10.1148/radiol.2019191586
- Payrovnaziri SN, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen JH, et al. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. J Am Med Inform Assoc. 2020 Jul 1;27(7):1173–85. https://doi.org/10.1093/jamia/ocaa053
- Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:14090473 [cs, stat] [Internet]. 2016 May 19; Available from: http://arxiv.org/abs/1409.0473
- Dong H, Suárez-Paniagua V, Whiteley W, Wu H. Explainable Automated Coding of Clinical Notes using Hierarchical Label-wise Attention Networks and Label Embedding Initialisation. Journal of Biomedical Informatics. 2021 Mar 9;103728. https://doi.org/10.1016/j.jbi.2021.103728
- Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research*. Journal of the American Medical Informatics Association. 2018 May 1;25(5):530–7. https://doi.org/10.1093/jamia/ocx160
- Mullenbach J, Wiegreffe S, Duke J, Sun J, Eisenstein J. Explainable Prediction of Medical Codes from Clinical Text. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans, Louisiana: Association for Computational Linguistics; 2018. p. 1101–11. Available from: https://www.aclweb.org/anthology/N18-1100
- Gorrell G, Song X, Roberts A. Bio-YODIE: A Named Entity Linking System for Biomedical Text. arXiv preprint arXiv:181104860. 2018;
- Bearryman E. Does Your Rare Disease Have a Code? 2016. https://www.eurordis.org/news/does-your-rare-disease-have-code
- Alex Ratner, Paroma Varma, Braden Hancock, Chris Ré, and other members of Hazy Lab. Weak Supervision: A New Programming Paradigm for Machine Learning [Internet]. SAIL Blog. 2019. Available from: http://ai.stanford.edu/blog/weak-supervision/

Thanks to members in Clinical Natural Language Processing Research Group and KnowLab
(please see the full list of members from the links)


Honghan Wu


Bea Alex


William Whiteley


Huayu Zhang


Víctor Suárez-Paniagua


Arlene Casey


Andreas Grivas


Minhong Wang


Emma Whitefield

Thank You & Questions

https://www.ed.ac.uk/usher/clinical-natural-language-processing
https://knowlab.github.io/

@EdiClinicalNLP