UNIVERSITY OF BIRMINGHAM

The patient voice as a form of clinical narrative: expanding ClinicalBERT with patient forum conversation

HealTAC 2021

Samantha Pendleton

Who am I?

- 3rd year PhD student in Clinical Bioinformatics
- I like ML clustering, NLP, semantics, and ontologies
- Current work on blood biomarkers and patient text on inflammation

- Patient unstructured text presents a rich but untapped resource for advancing patient care
- Clinicians can only capture so much from a patient - there could be further information contained in social media

UNIVERSITY OF BIRMINGHAM

Patient Voice

- How patients express their conditions
    - In person appointments or online via social media, blogs, forums
- There is growing recognition of the importance of the patient voice in research and clinical practice
- The patient voice can be a secondary resource for research

What is an ontology

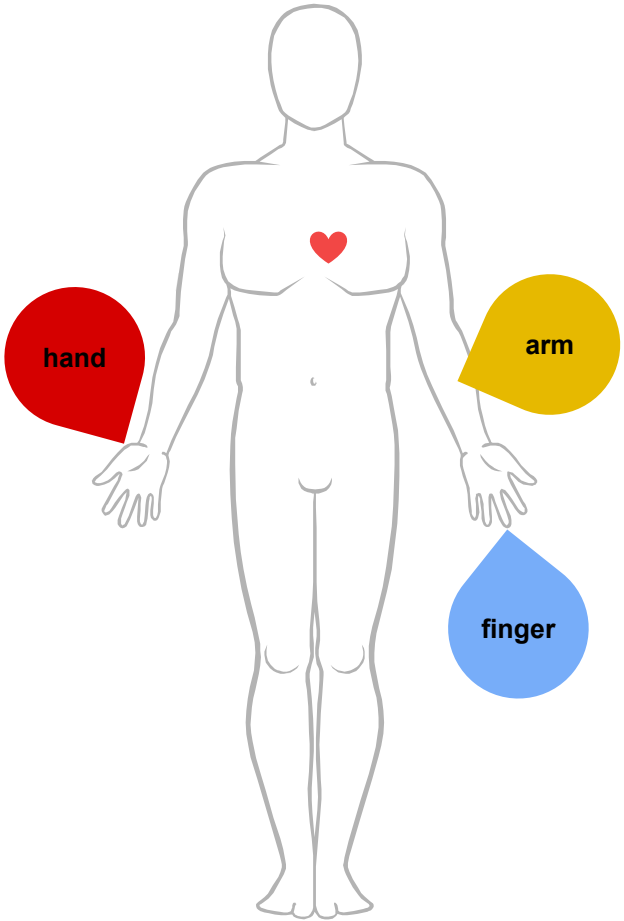
Represents a domain of knowledge - a THING

Concepts (e.g. anatomy)

Classes	**finger**, **hand**, **arm**		
Relationships	**finger**	*partOf some*	**hand**
	hand	*partOf some*	**arm**
Axioms	**finger**	*partOf some*	**arm**

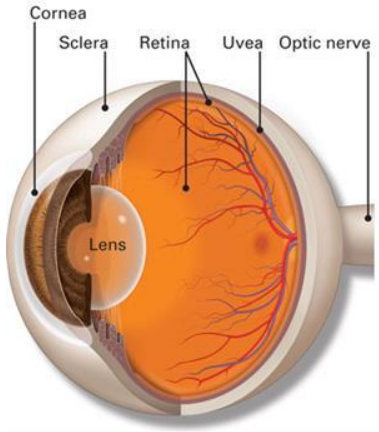
Annotations/Metadata

definitions	(e.g. arm: “front limbs of an animal”)
x-refs	(cross-references to other ontologies)
synonyms	(e.g. forelimb = arm)

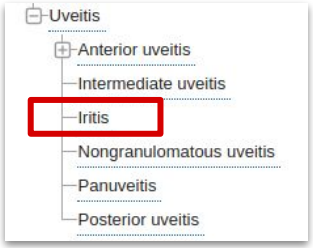


Previous work

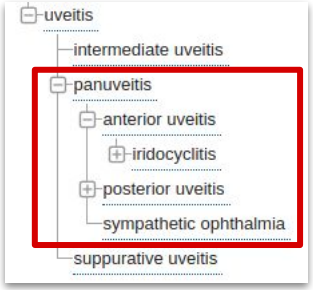
- Created an ontology, **Ocular Immune-Mediated Inflammatory Diseases Ontology (OcIMIDo)**
- Started by looking at uveitis, inflammatory eye disorder in the uvea, which wasn't systematically characterised in current biomedical ontologies
    - These ontologies are not fully representing our domain of interest
- Lack of characterisation of uveitis phenotypes to distinguish patient cohorts



HPO

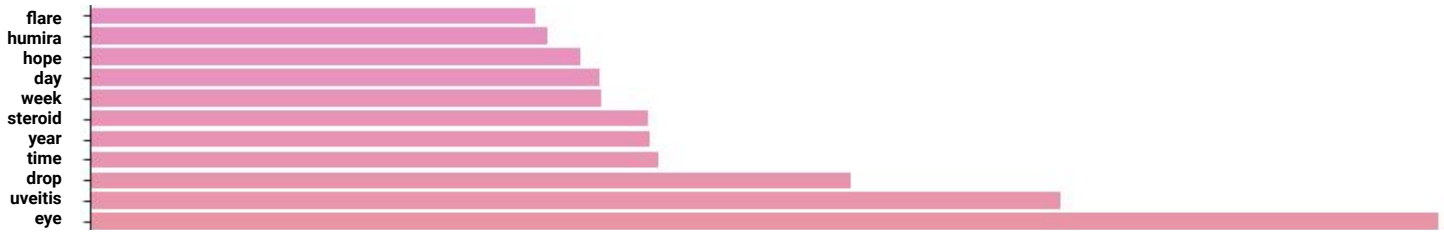


DOID



OcIMIDo ontology

- We developed OcIMIDo with expert domain knowledge and existing clinical guidelines
- We expanded the ontology with patient-preferred terms from patient forum conversation via a novel synonym extraction method (with tf-idf)
- We saw that when using patient-preferred terms alongside clinical terms, our text mining efforts were increased
- As an application we performed sentiment analysis

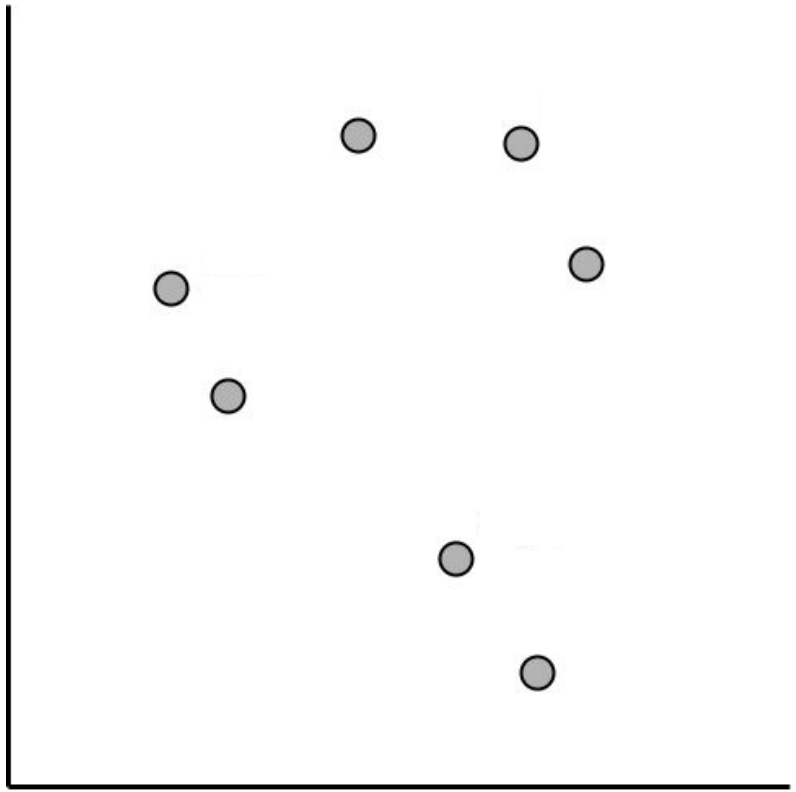


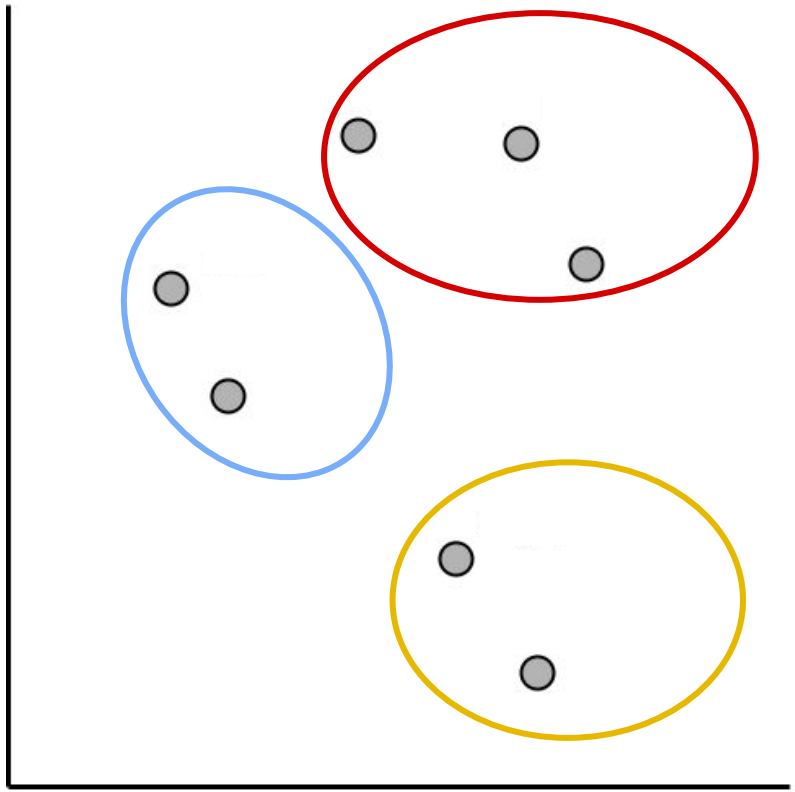
Expanding **clinical letters** with **patient forum** conversation

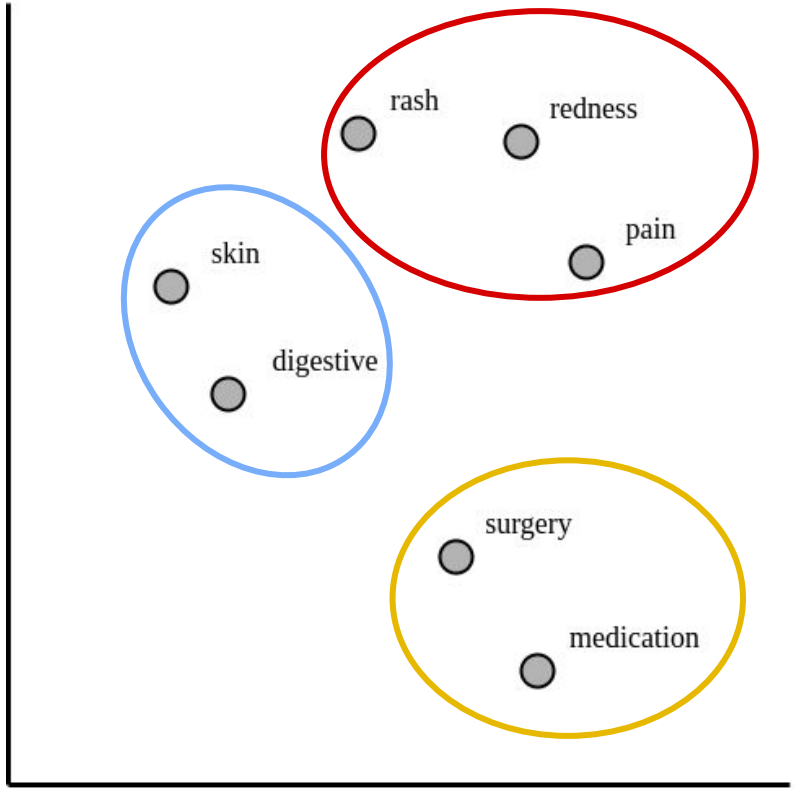
What are clinical letters?

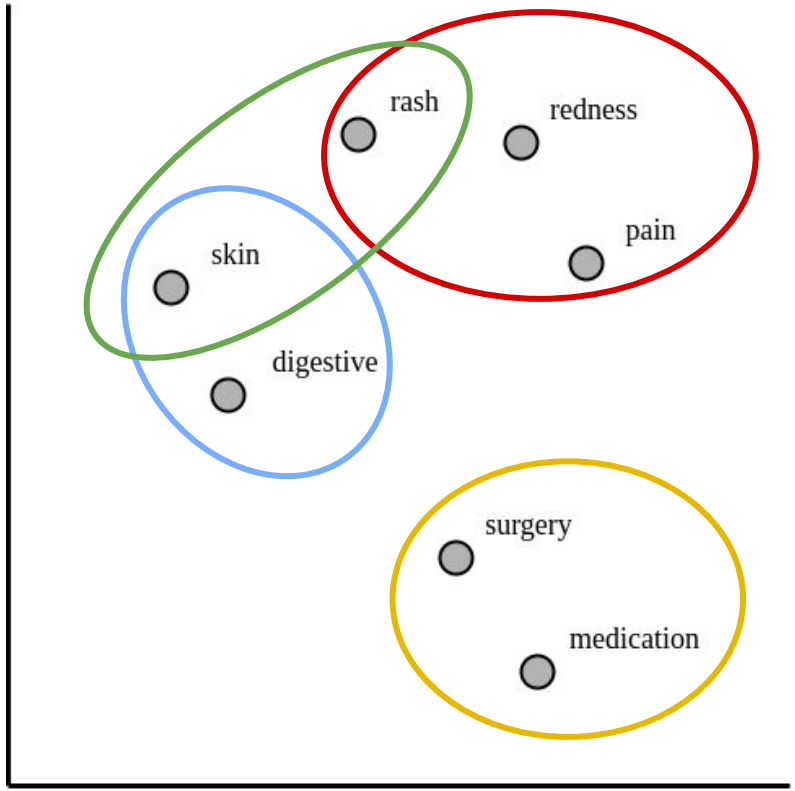
- Clinical letters typically describe a patient's phenotype
    - Symptom description, test performed, and diagnosis
- MIMIC-III is a large database, freely available, consisting of 40,000 critical care patients between 2001 and 2012
- ClinicalBERT contextually represent 100,000 MIMIC-III clinical letters via embeddings
- Embeddings learn word associations (semantic similarity): another method of synonym curation

UNIVERSITY OF BIRMINGHAM






rash
redness
pain
skin
digestive
surgery
medication


rash
redness
pain
skin
digestive
surgery
medication

Why is ClinicalBERT interesting?

- ClinicalBERT can be a different narrative for conditions (compared to already existing publications)
    - We can find word associations through clinical letters
- Expanding ClinicalBERT with patient forum conversation can also be a different narrative
    - Another perspective into patient-preferred terms, which we've already shown the usefulness for text mining efforts

Why is ClinicalBERT interesting? (part 2)

- Hypothesis
  - Can ClinicalBERT be expanded with forum conversation and be a good secondary resource for data analysis? If so...
  - Capturing patient voice through embeddings - what can we uncover?

Workflow

1

Built a basic **U**nified In**f**lammatory **O**ntology (**UFO**) based on ICD-10 structure

Includes synonyms from other biomedical ontologies and layperson terms

2

Extracted 67,204 individual inflammatory-related posts from patient forum conversations

Such as: appendicitis, blepharitis, cystitis, folliculitis, rhinitis, and more...

3

Expanded ClinicalBERT (MIMIC-III 100,000 clinical letters) with the patient forum conversation

Observed some initial differences (statistically significant)

4

Curated synonyms from both embedding models and investigated how well both models capture synonyms

Brief observations

- Frequent misspellings in clinical letters, even more in the forum
- Abbreviations used in frequently in clinical letters and the forum

ClinicalBERT

- **Fasciitis**: fascitis & faciitis & fasciatis & fascititis & fasc

Forum

- **Fasciitis**: faciitis & fascitis & fascititis & faciatis & more misspellings & PF

Top-5 observations

- High semantic similarity terms (top-5) for ClinicalBERT were often other conditions

Cystitis					
ClinicalBERT	prostatitis	proctitis	hematuria	pyelonephritis	enteritis
Forum	prostatitis	UTI	UTIs	pyelonephritis	IC
Sinusitis					
ClinicalBERT	mastoiditis	tracheitis	pyelonephritis	epiglottitis	bronchitis
Forum	sinusitus	sinus	pancreatitus	pancreaitis	pancreatitis

KEY: new term / misspellings / abbreviations

What does this mean?

- ClinicalBERT top-5 being other conditions could mean many things:
  - These other conditions are being tested - e.g. prostatitis is a condition in men that is similar to cystitis

- Forum top-5 conditions being different from ClinicalBERT could mean:
  - Comorbidities / non-common associations - e.g. pancreatitis and sinusitis
  - A google search...

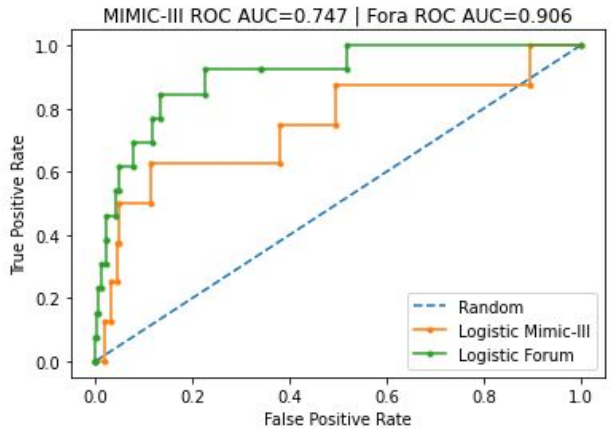
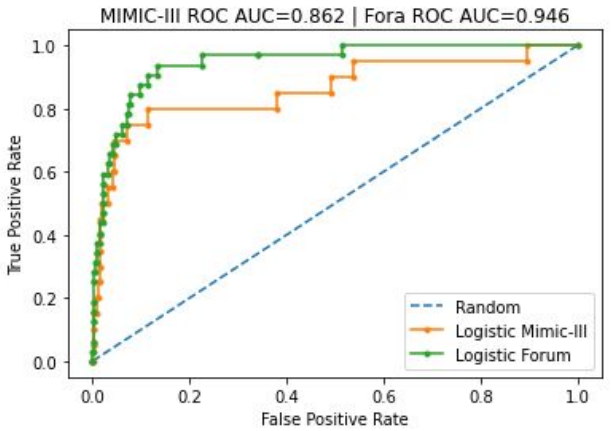
Comment	I know this is going to sound weird, but when my pancreatitis flares up, I seem to have sinus problems. Like one of my sinuses will swell shut.
Reply	My pain often starts in the throat and esophagus then moves down into my pancreas. No idea what this is and neither do Doctors.

Statistics

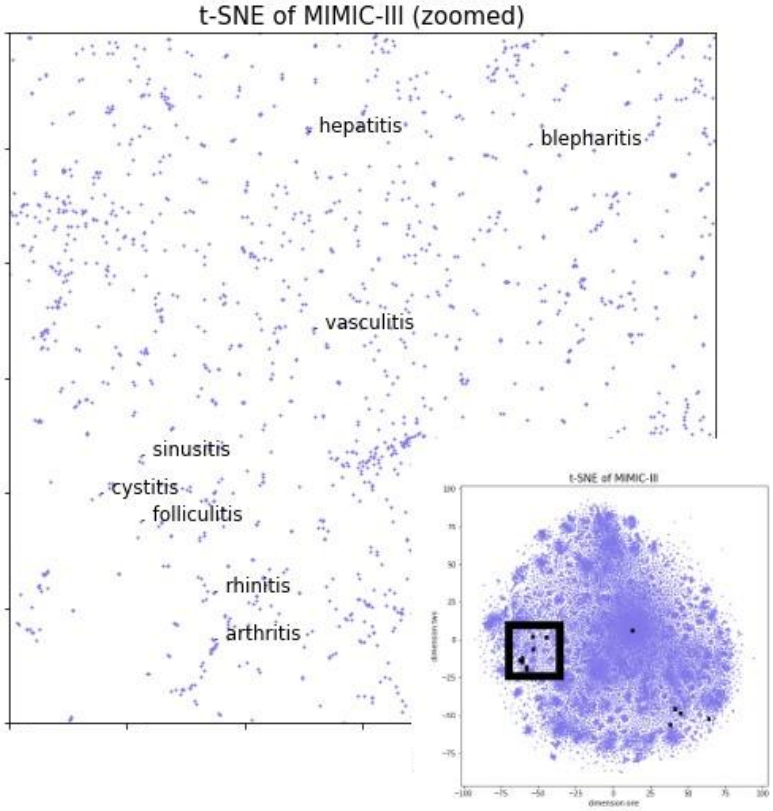
- Wilcoxon signed-rank paired-test: observing how semantic similarity scores changed in some inflammatory terms of interest with expansion
- Majority terms statistically significant...except:
    - **blepharitis -** perhaps because blepharitis not an intensive care emergency? Or do clinicians and patients talk about blepharitis in the same way?

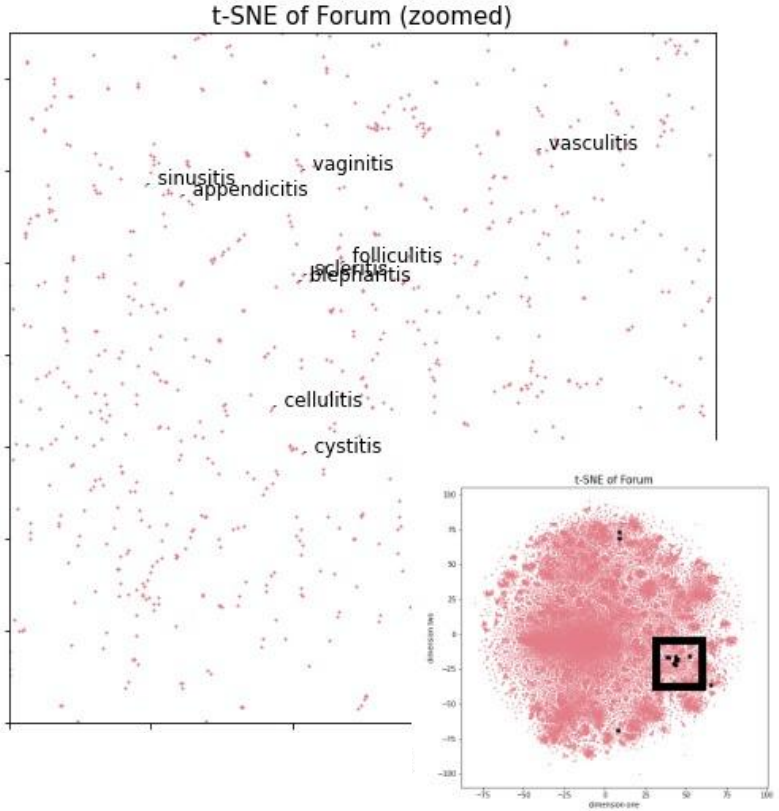
ROC AUC

- When looking into inflammatory terms of interest and their synonyms:

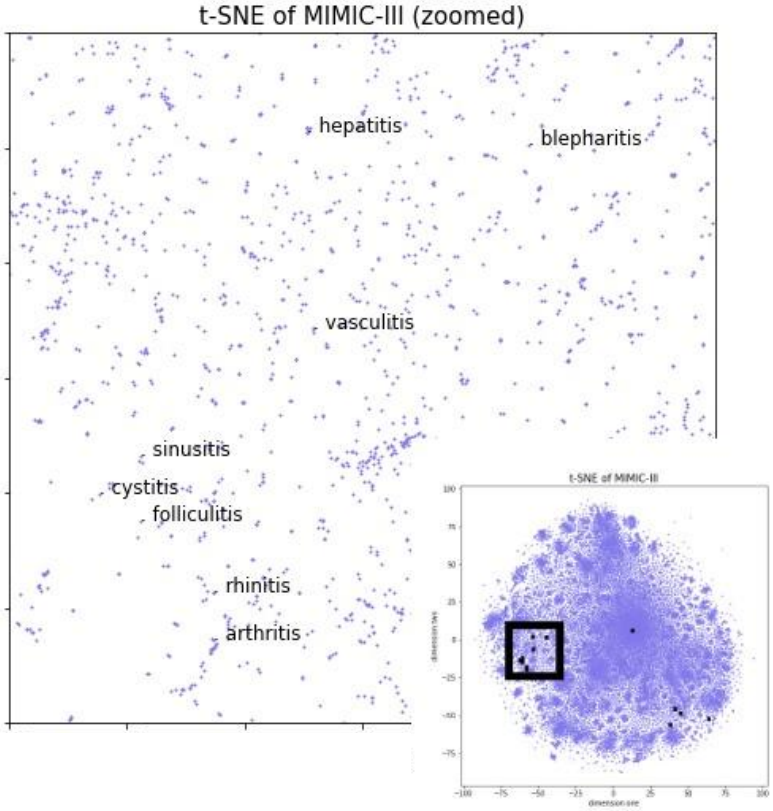
Using UFO synonyms	Expansion synonym from embeddings
	

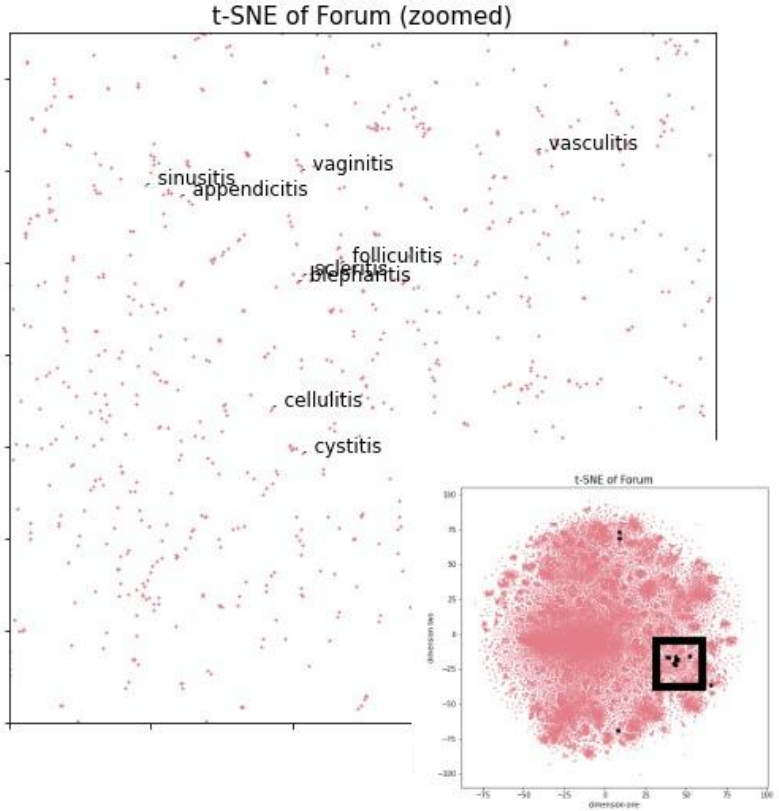
t-SNE plots of embeddings



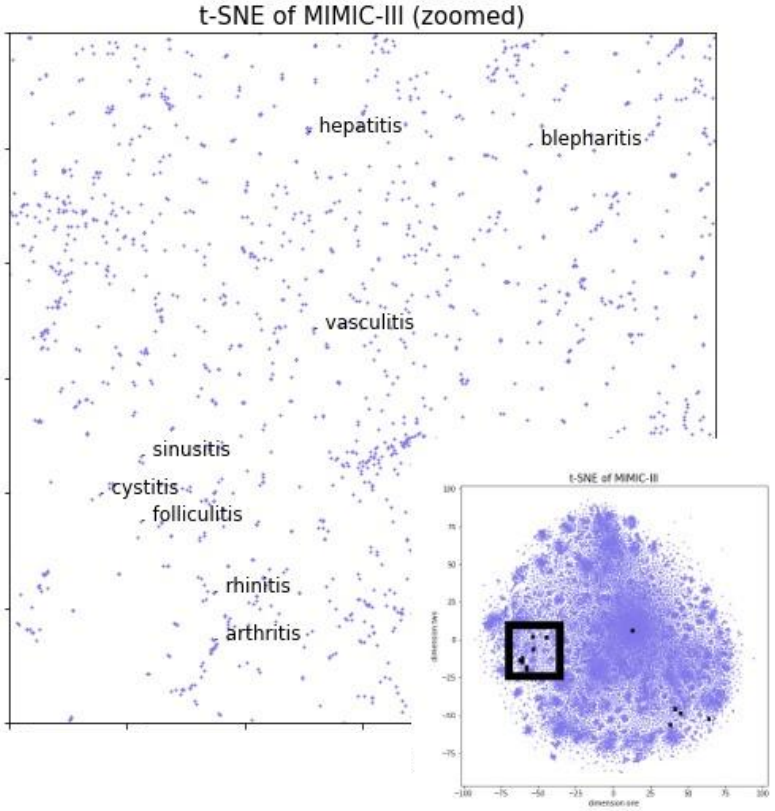


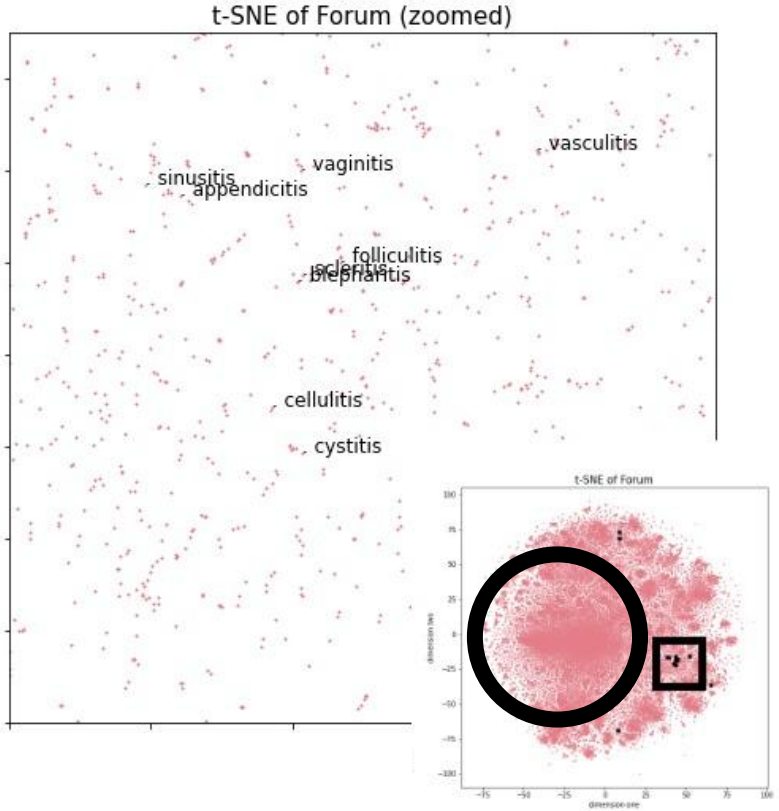
1. Coordinates of conditions



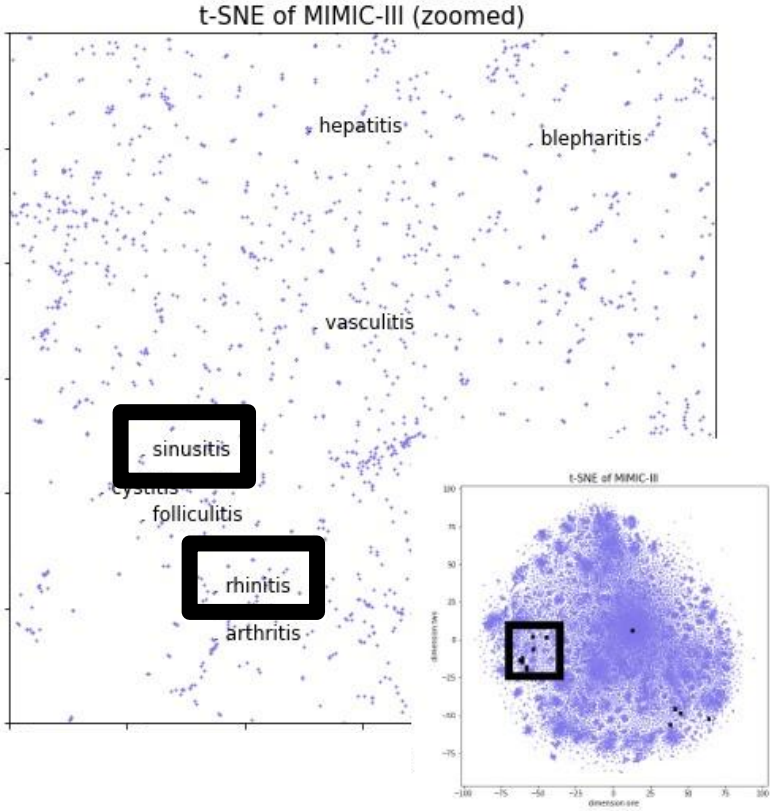


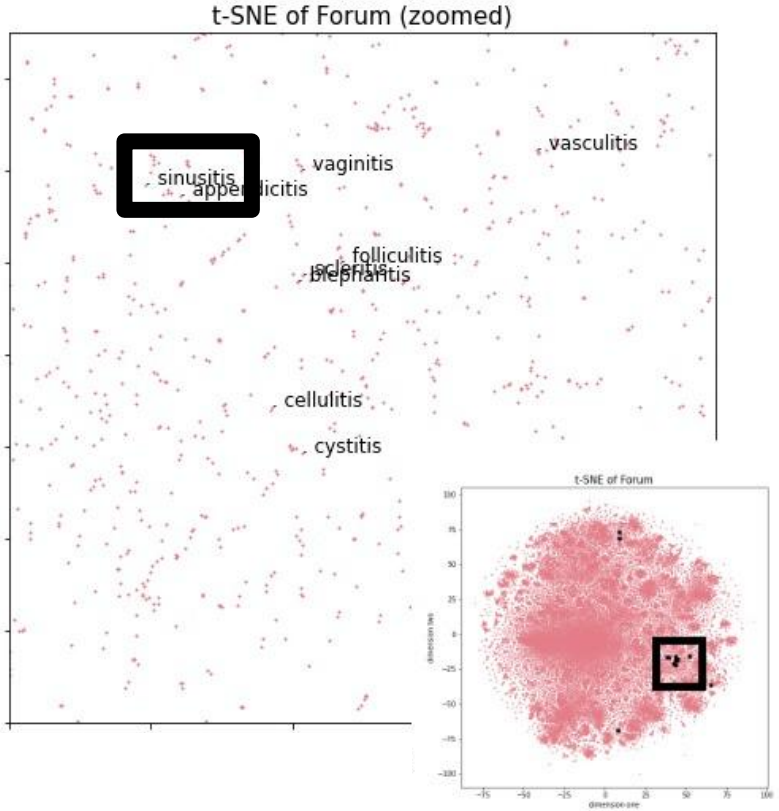
2. Looking at shapes



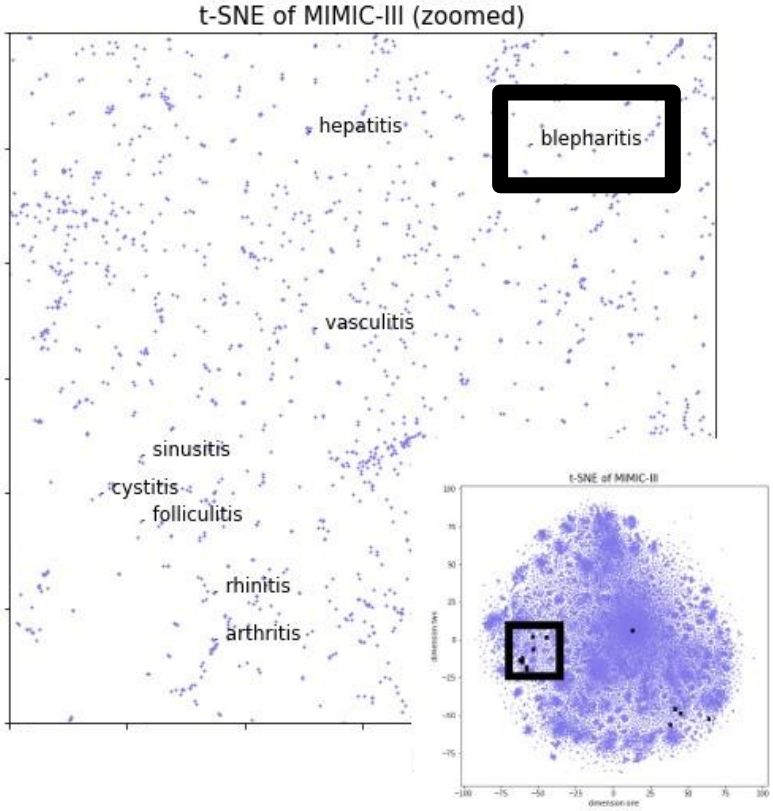


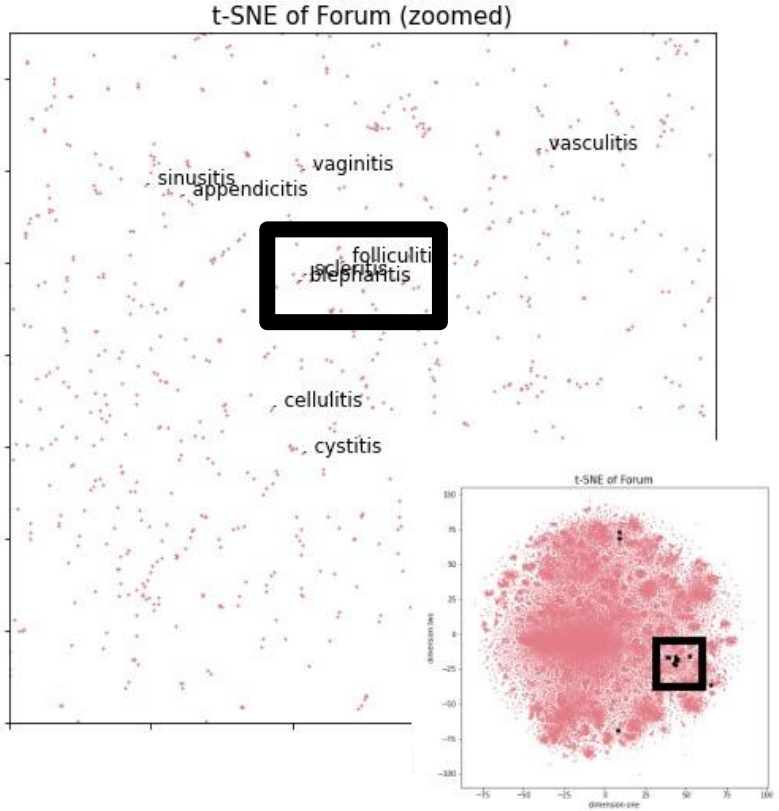
3. Respiratory: rhinitis and sinusitis





4. Visual: blepharitis & scleritis





Discussion

- We observed clinicians & patients make spelling mistakes & use abbreviations
- We identified new synonyms from ClinicalBERT and the forum model
- The forum embeddings were better at capturing synonyms and synonyms were more frequent in the forum - showing us clinical letters don’t usually consist of synonyms
- Observed differences in the way conditions are discussed in the different contexts

Conclusion

- We introduced the patient voice as a secondary resource for research
- Study demonstrated that the patient’s voice is a valuable resource as a clinical narrative
- Presented possible novel associations in the forum

Future Endeavours

- Clustering to observe groups and to characterise
- Investigate further into the forum model to uncover these possible novel associations
- These associations may give greater insight into medical conditions and better informing clinical care

UNIVERSITY OF BIRMINGHAM

Thank you!

Samantha Pendleton

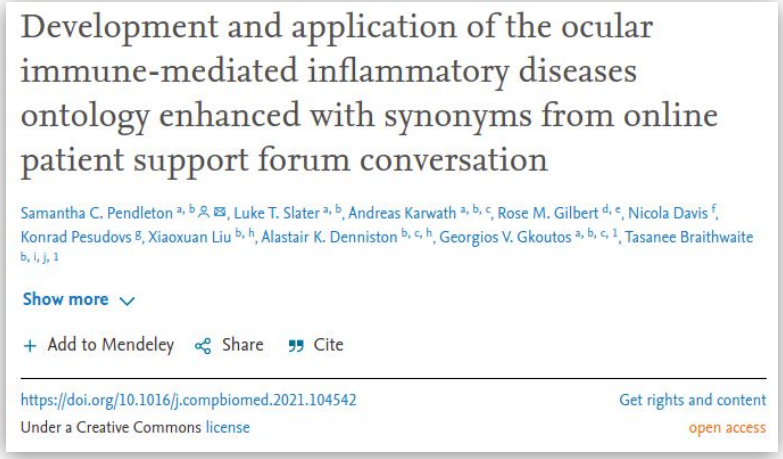
Email me to ask further questions!

scp887@student.bham.ac.uk

Acknowledgements

- George Gkoutos
- Luke Slater & Andreas Karwath
- Gkoutos group, University of Birmingham

- Tasanee Braithwaite
- Members at Moorfields Eye Hospital
- Members at Institute of Inflammation and Ageing, University of Birmingham
- Members at Olivia's Vision



Development and application of the ocular immune-mediated inflammatory diseases ontology enhanced with synonyms from online patient support forum conversation

Samantha C. Pendleton [a, b], Luke T. Slater [a, b], Andreas Karwath [a, b, c], Rose M. Gilbert [d, e], Nicola Davis [f], Konrad Pesudovs [g], Xiaoxuan Liu [b, h], Alastair K. Denniston [b, c, h], Georgios V. Gkoutos [a, b, c, 1], Tasanee Braithwaite [b, i, j, 1]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.compbiomed.2021.104542

Get rights and content

Under a Creative Commons license

open access

OcIMIDo Paper: bit.ly/3iJXT3g

UNIVERSITY OF BIRMINGHAM

MRC Medical Research Council

@sap218