

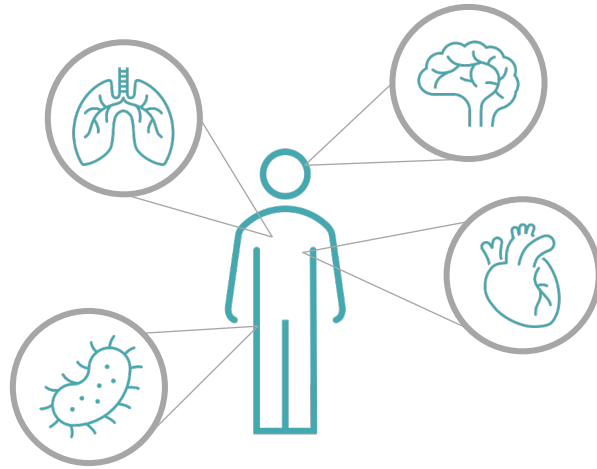
Co-morbidity Representation in Artificial Intelligence: Tapping into Unused Clinical Knowledge

William Bolton

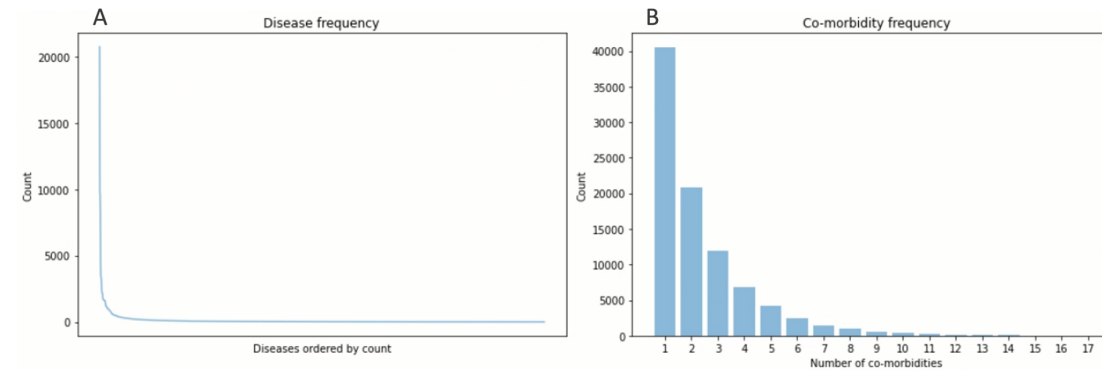
W3PHIAI-24

27th February 2024

Diagnosis information can be hard to apply to AI systems.



Co-morbidities or chronic long-term medical conditions are a **major challenge in healthcare**



Challenges such as **combinatorial complexity**, **heterogeneity**, and a **lack of data** make using disease data in AI systems difficult

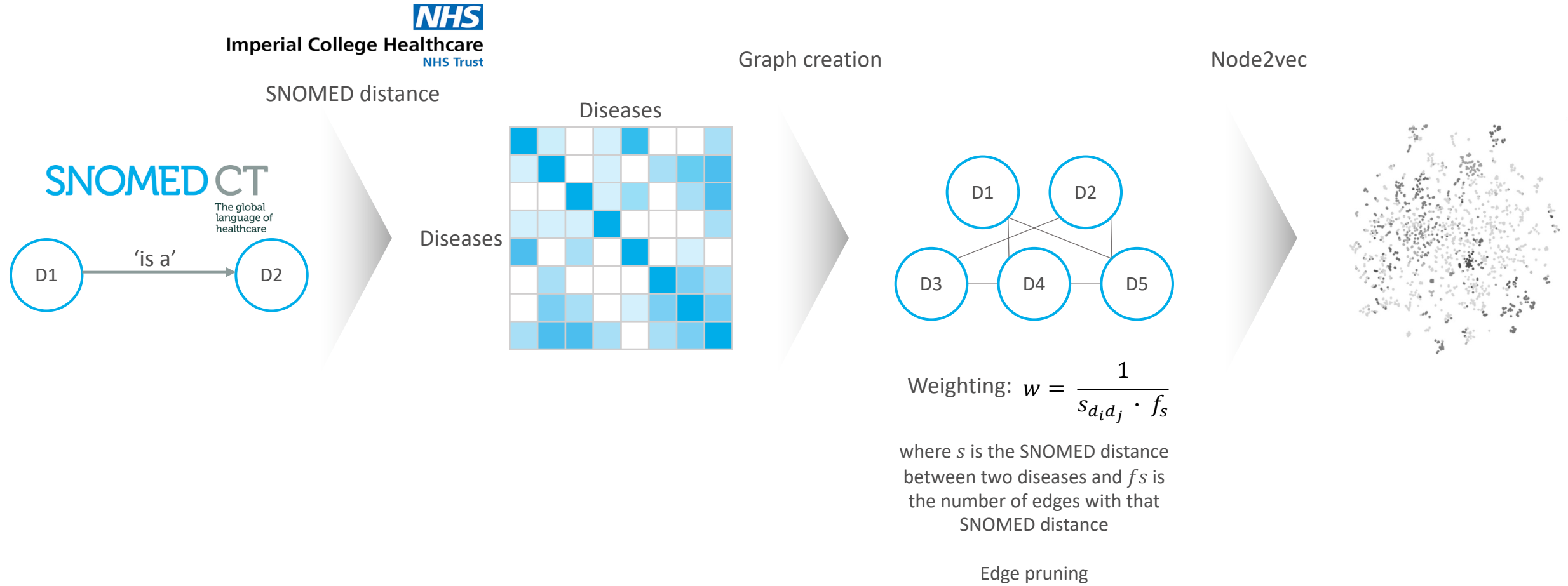


Existing AI research on co-morbid patients does not tackle these problems and therefore **lacks appropriate representation**

Aim

Creating **meaningful embeddings** from **external medically grounded knowledge**, to help **overcome such challenges** and **support downstream AI applications**

Our pipeline leverages a publicly available expert curated healthcare knowledge graph.



Optimized for the mean SNOMED distance between each disease and their nearest neighbor

We tested our methodologies against two clinically relevant AI tasks.

Supervised learning task

SNOMED disease embeddings for each patient



In-hospital mortality within a year

Long length of stay

- Model: Set transformer

- Baseline:

Logistic regression with one hot encodings

Logistic regression with Charlson co-morbidity categories

Set transformer with random SNOMED disease embeddings

- Evaluation: AUROC

Similar patient retrieval task

- Useful for **case-based reasoning**
- Created **co-morbid patient embeddings** as the **mean** of all the patients SNOMED disease embeddings
- Retrieved similar patients through **nearest neighbor lookup**

- Baseline:

One hot encodings

Rocheteau metric

- Evaluation:

Metrics

SNOMED similarity score

Charlson Jaccard index

Expert humans

Patient co-morbidity similarity questionnaire

Thank you for taking the time to complete our survey. For each question please select the patient (A, B or C) you believe is the most similar to the patient in question with regards to their co-morbidities.

Please note that these co-morbidities were extracted from SNOMED codes and may be historic diagnosis.

[Sign in to Google to save your progress. Learn more](#)

* Indicates required question

Patient in question has: Hypertensive disorder, Hypercholesterolemia and Neoplasm of kidney

Patient A: Hypertensive disorder, Neoplasm of kidney

Patient B: Hypertensive disorder, Hypercholesterolemia, Insulin treated type 2 diabetes mellitus, Neoplasm of kidney

Patient C: Hypertensive disorder, Diabetes mellitus type 2, Hypercholesterolemia, Osteoarthritis, Atrial fibrillation, Mixed anxiety and depressive disorder, Malignant tumor of prostate, Neoplasm of kidney

Any comments or questions on this case?

Your answer _____

Two novel metrics were created for the Similar patient retrieval task.

SNOMED similarity score

$$SNOMED\ sim_{p1,p2} = f(S_{p1,p2}) + f(S_{p2,p1})$$

where $S_{p1,p2}$ is a SNOMED distance matrix for the patients co-morbidities

We match each disease of $p1$ to a disease of $p2$ so that the matching minimized the following equation:

$$f(A) = \sum_{i=1}^n \min_{j \in \{1, \dots, m\}} \left(1 - \frac{1}{A_{ij} + 1} \right)$$

where $A \in \mathbb{R}^{n \times m}$

More similar patients return **smaller** values

Charlson Jaccard index

The Charlson co-morbidity index is a widely adopted clinical tool that classifies some specific co-morbidities to 17 different categories

$$Charlson\ Jaccard\ index_{p1,p2} = \frac{|C_{p1} \cap C_{p2}|}{|C_{p1} \cup C_{p2}|}$$

where C represents the set of Charlson co-morbidities for a particular patient

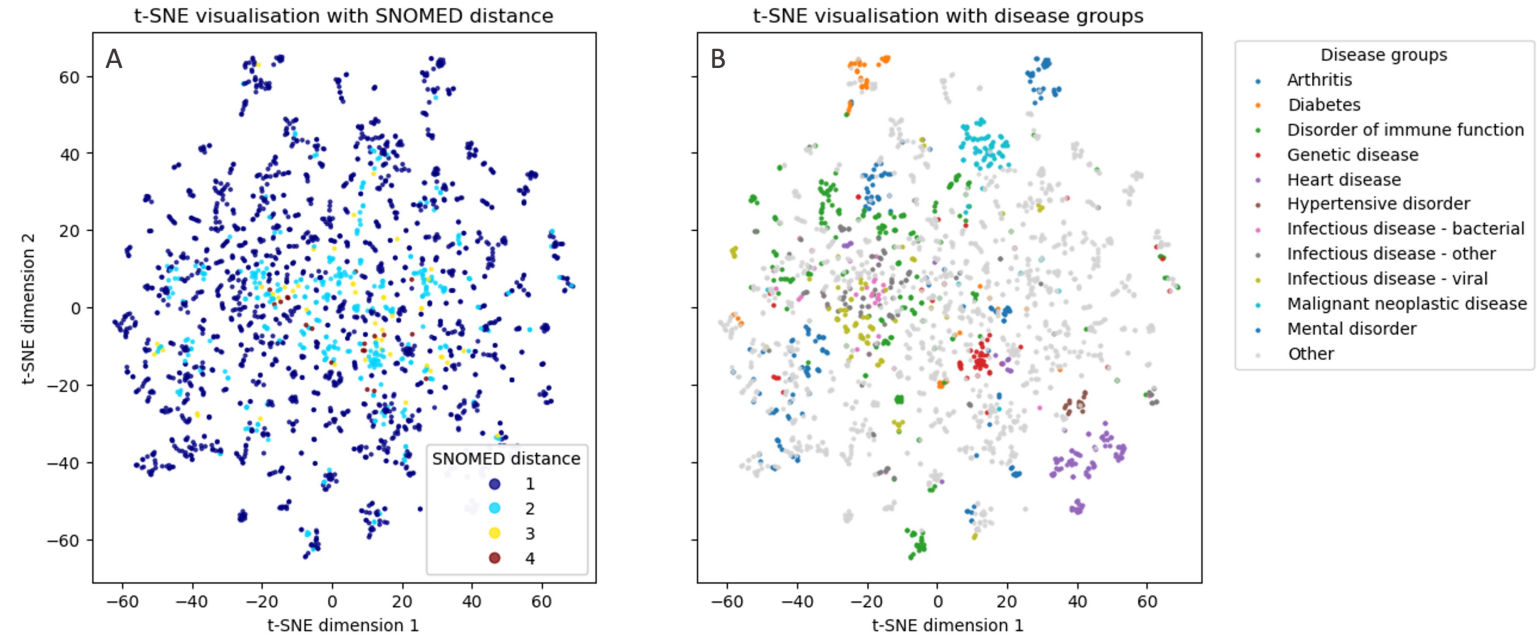
More similar patients return **larger** values

Our pipeline generated meaningful SNOMED disease embeddings.

Imperial College Healthcare
NHS Trust

95,157 patients

2,133 chronic conditions



- **Optimization of pre-processing** returned a mean SNOMED distance between each node and their nearest neighbor of **1.23**
- **Related conditions are located close in the embedding space**
- For example:
 - Viral and bacterial infectious diseases
 - Heart diseases and hypertensive disorders

SNOMED disease embeddings are informative features for models undertaking clinically relevant predictions.

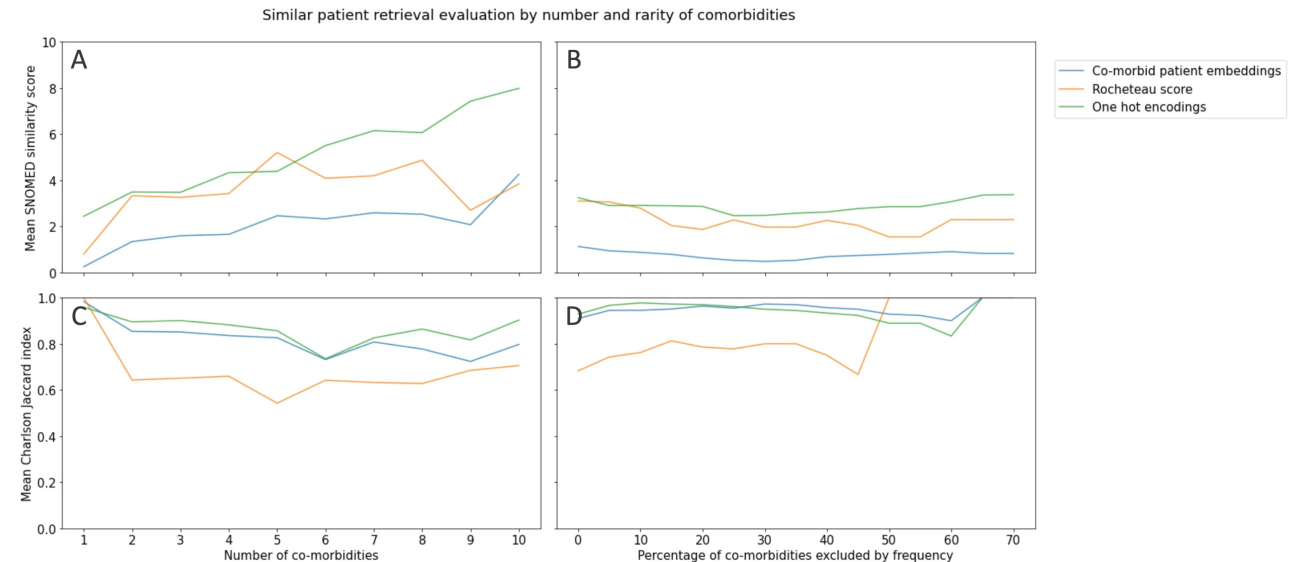
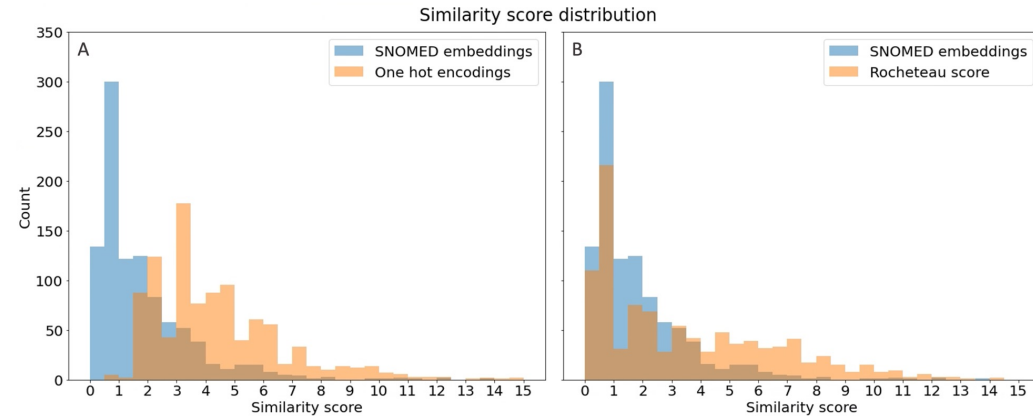
Table 1: Mean unseen test set AUROC results for supervised learning classification tasks in different populations.

Features	Model	Year Mortality		Long length of stay	
		Overall	Rarest co-morbidities	Overall	Rarest co-morbidities
Charlson co-moribidity categories	Logistic regression	0.65 (SD 0.01)	0.50 (SD <0.01)	0.60 (SD 0.01)	0.50 (SD 0.03)
One hot encodings	Logistic regression	0.79 (SD 0.02)	0.80 (SD 0.23)	0.72 (SD 0.01)	0.55 (SD 0.11)
Random SNOMED disease embeddings	Set transformer	0.80 (SD 0.03)	0.56 (SD 0.33)	0.74(SD 0.02)	0.52 (SD 0.23)
SNOMED disease embeddings	Set transformer	0.82 (SD 0.02)	0.85 (SD 0.14)	0.75 (SD 0.01)	0.61 (SD 0.20)

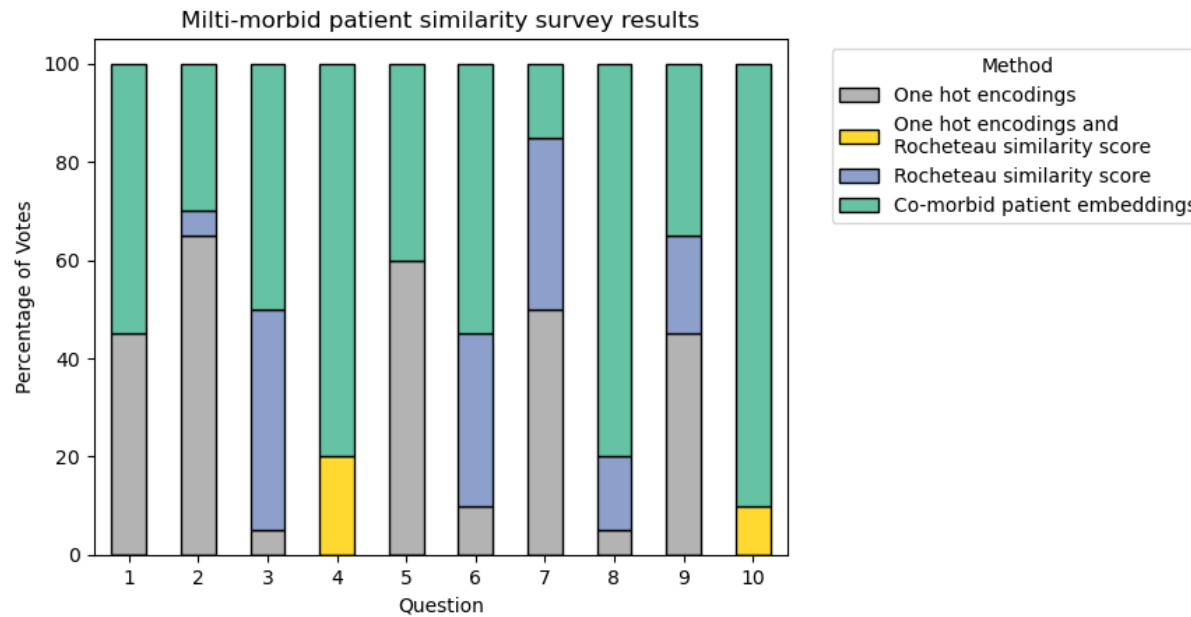
Co-morbid patient embeddings finds more similar patients and is consistent across different degrees of rarity and multi-morbidity.

Table 2: Mean results for the similar patient retrieval task.

Method	SNOMED similarity score	Charlson Jaccard index
One hot encodings	4.40 (SD 2.32)	0.88 (SD 0.30)
Rocheteau's method	3.52 (SD 3.26)	0.69 (SD 0.20)
Co-morbid patient embeddings	1.78 (SD 1.90)	0.84 (SD 0.34)



Patients identified by our method were selected as the most similar in 60% of questions with a mean winning margin of 40%.



	Co-morbidities			
Question 8 patient	Gestational diabetes mellitus	Hypertensive disorder	Pre-eclampsia	Varicella
Co-morbid patient embeddings	Gestational diabetes mellitus	Pregnancy-induced hypertension	Pre-eclampsia	Varicella
Rocheteau score	Gestational diabetes mellitus	Hypertensive disorder	-	Varicella
One hot encodings	Gestational diabetes mellitus	-	Pre-eclampsia	Varicella
Question 10 patient	Osteoarthritis	Alcoholism		
Co-morbid patient embeddings	Osteoarthritis	Alcohol dependence		
Rocheteau score	Osteoarthritis	Alcoholism	Peripheral nerve entrapment	
One hot encodings	Osteoarthritis	Alcoholism	Peripheral nerve entrapment	

Legend for similarity:

- Identical (Green)
- Similar (Yellow)
- Dissimilar (Red)

Co-morbidity Representation in Artificial Intelligence: Tapping into Unused Clinical Knowledge.

Conclusion

- Developed a **novel pipeline** to extract and utilize the previously untapped information contained within SNOMED CT to **represent diseases and co-morbid patients**
- Demonstrated its **utility** in **classification** and **similar patient retrieval tasks** with automatic and human evaluation techniques
- Our approach is **generalizable** and able to **overcome many of the problems** associated with using disease data in AI
- We hope this research can **progress AI systems** towards **clinical benefit**

Thank you!

William Bolton

W3PHIAI-24

27th February 2024

william.bolton@imperial.ac.uk

Linked 



**Imperial College
London**



GitHub

