Semantic Enrichment of XAI explanation for Healthcare

Luca Corbucci
Ph.D. Candidate - Computer Science Department
📧 **luca.corbucci@phd.unipi.it**
Supervisor: Anna Monreale

👋 Hi, I'm Luca

- 🧑‍💻 Ph.D. Student @ University of Pisa
- 🎤 Podcaster @ PointerPodcast
- 🦸 Community Manager @ SuperHeroesValley
- 🌏 lucacorbucci.me

Outline

- Introduction
- Related Works
- Goal of the project
- Methodology
- Conclusions

Introduction

Introduction

- Deep Learning is replacing the classic artificial neural network techniques because of their better performances if high-dimensional datasets are available.

- The most significant drawbacks of Deep Learning models which hold back the use in the real world is their black-box nature

- These systems hide their internal logic to the user and even the developers do not know how they have reached their conclusions.

Explainable AI

The goal is to “open the black boxes”
to build a more explainable, trustworthy
and ethical machine learning

Why do we need an explanation?

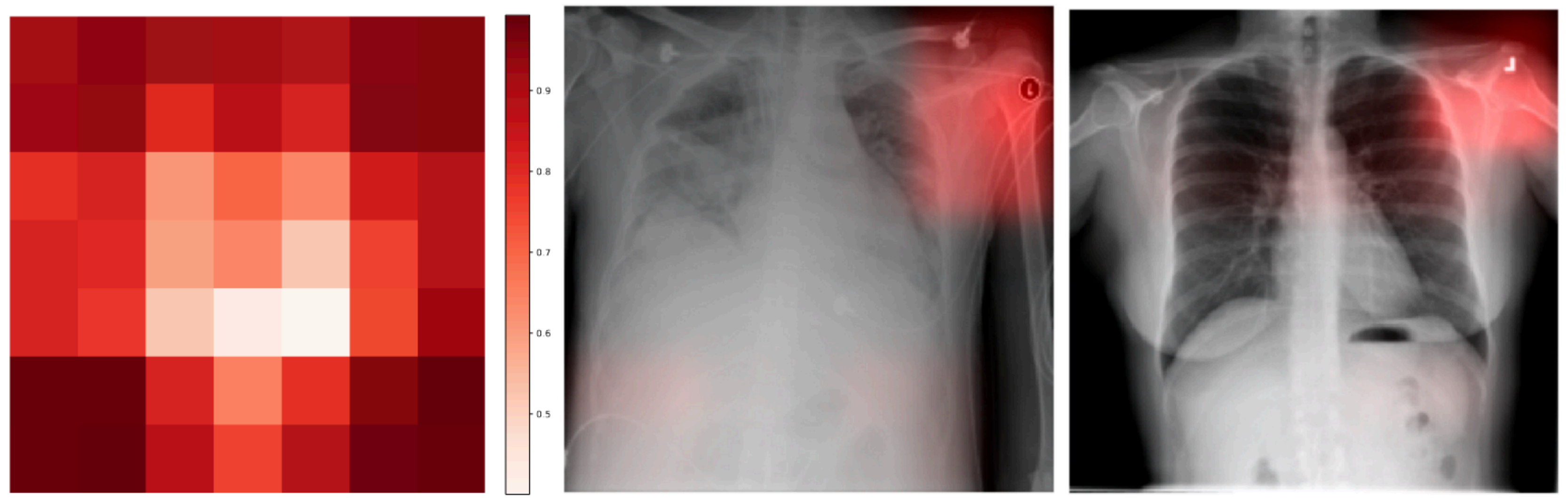
- To discover biases in a model
- To understand why a certain decision was made and to increase the trust in the model
- To avoid a right prediction for the wrong reason
- To be sure that a model will work even if I switch my equipment
- It is a legal requirement prescribed by Art. 22 of the GDPR

Examples:

To discover biases in a model

COMPAS: a model to predict the risk of criminal recidivism. It was found [1] to have an ethnic bias:

To be sure that a model will work even if I switch my equipment



The predictions made by a CNN using x-rays image were found to be influenced by “Confounding variables” [2]

To avoid a right prediction for the wrong reason

(a) Husky classified as wolf (b) Explanation

A model has been trained to recognize wolves and husky dogs, the black box was making its predictions to classify a wolf solely on the presence of snow in the background. [3]

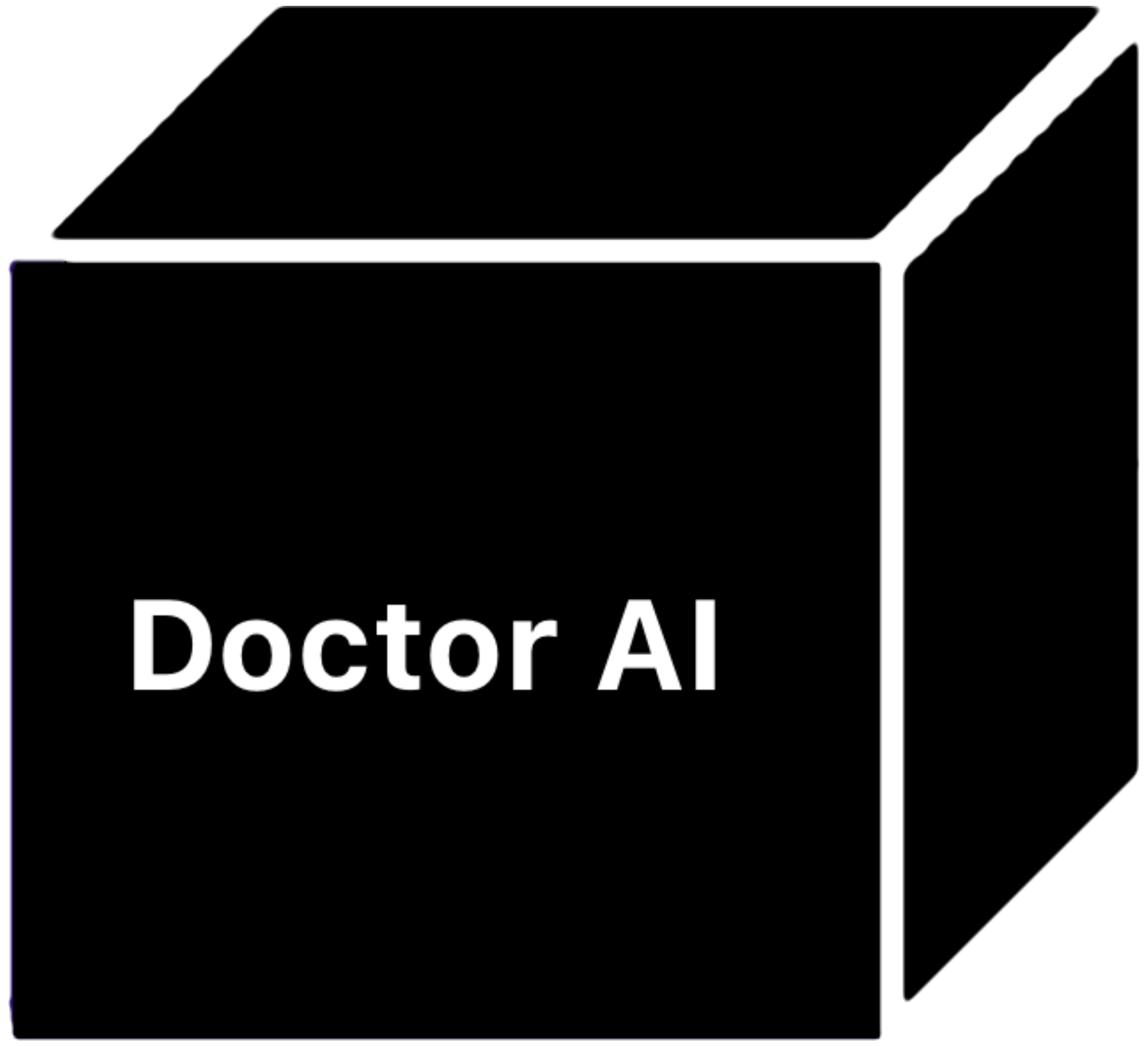
[1] How We Analyzed the COMPAS Recidivism Algorithm - https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
[2] Confounding variables can degrade generalization performance of radiological deep learning models - Zech, John R. and Badgeley, Marcus A. and Liu, Manway and Costa, Anthony B. and Titano, Joseph J. and Oermann, Eric K.
[3] "Why Should I Trust You?": Explaining the Predictions of Any Classifier - Ribeiro, Marco Tulio and Singh, Sameer and Guestrin, Carlos

Related works

What is Doctor AI?

INPUT:
Clinical History of a patient

1) ['562.12', '280.0', '211.3', '401.9', '250.00', '702.19', 'V10.3']
2) ['562.12', '276.0','250.00', '401.9', 'V10.3']
3) ['584.9', '276.5', '585', '532.90', '250.00', '285.9', 'V10.3', 'V44.2']
4) ['569.69', '560.89', '998.59', '038.9', '995.91', '584.9', '585.9', '998.32', '250.00']



OUTPUT:
Predictions of future diseases

Doctor AI: Predicting Clinical Events via Recurrent Neural Networks -
Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, Jimeng Sun

Can we explain the reason behind a prediction?

What is Doctor XAI?

Given the instance we want to explain we search for the most similar ones in the dataset

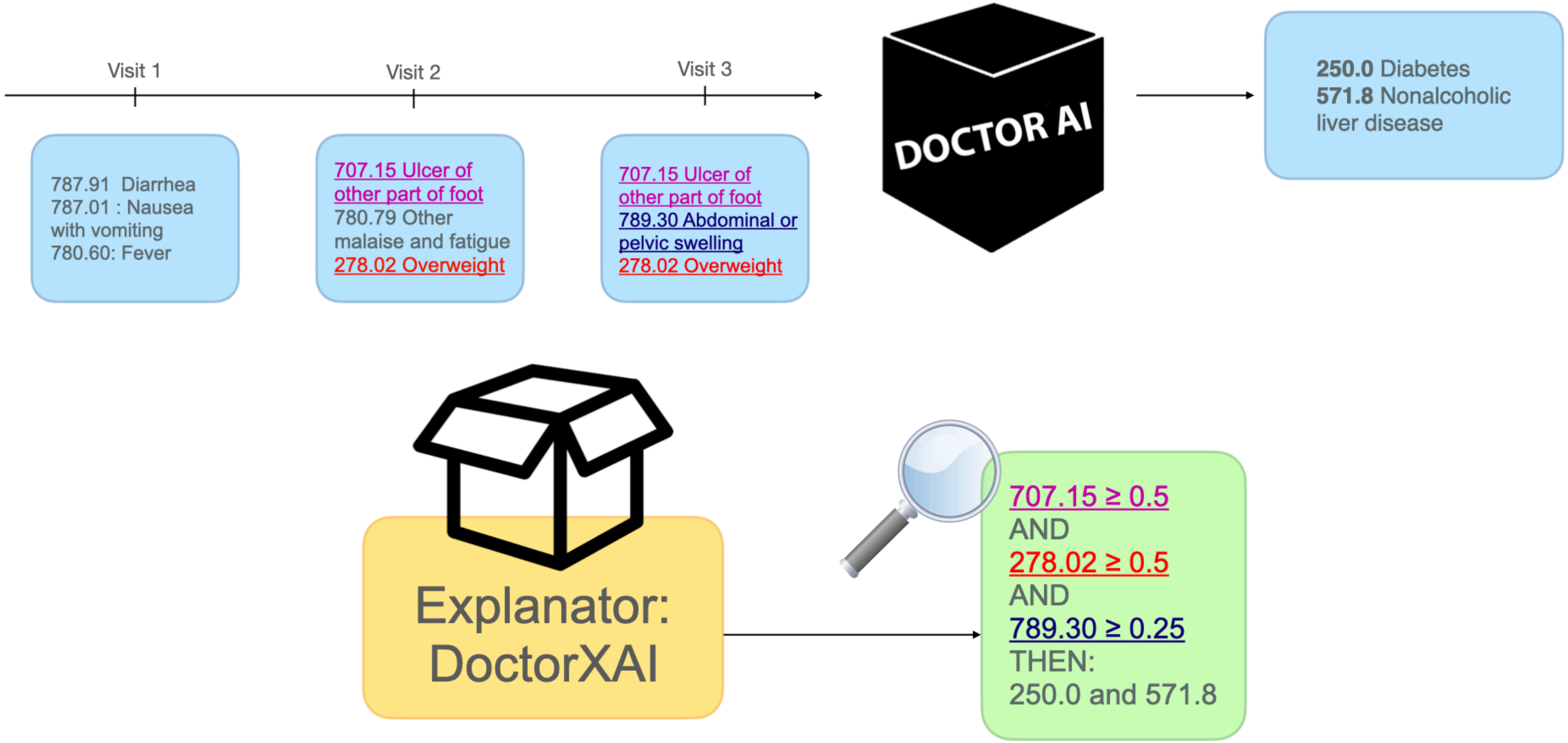
Some synthetic instances Are generated and Classified using Doctor AI

The data given as input To Doctor AI and the Corresponding output Is used to train a Decision tree

Doctor XAI returns the rule That led to the Doctor AI prediction

Doctor XAI: an ontology-based approach to black-box sequential data classification explanations - Panigutti, C, Perotti, A, Pedreschi, D

How Doctor XAI works



Goal of the project

What have we done?

Doctor AI

It makes predictions
Using clinical
Data of a patient

Doctor XAI

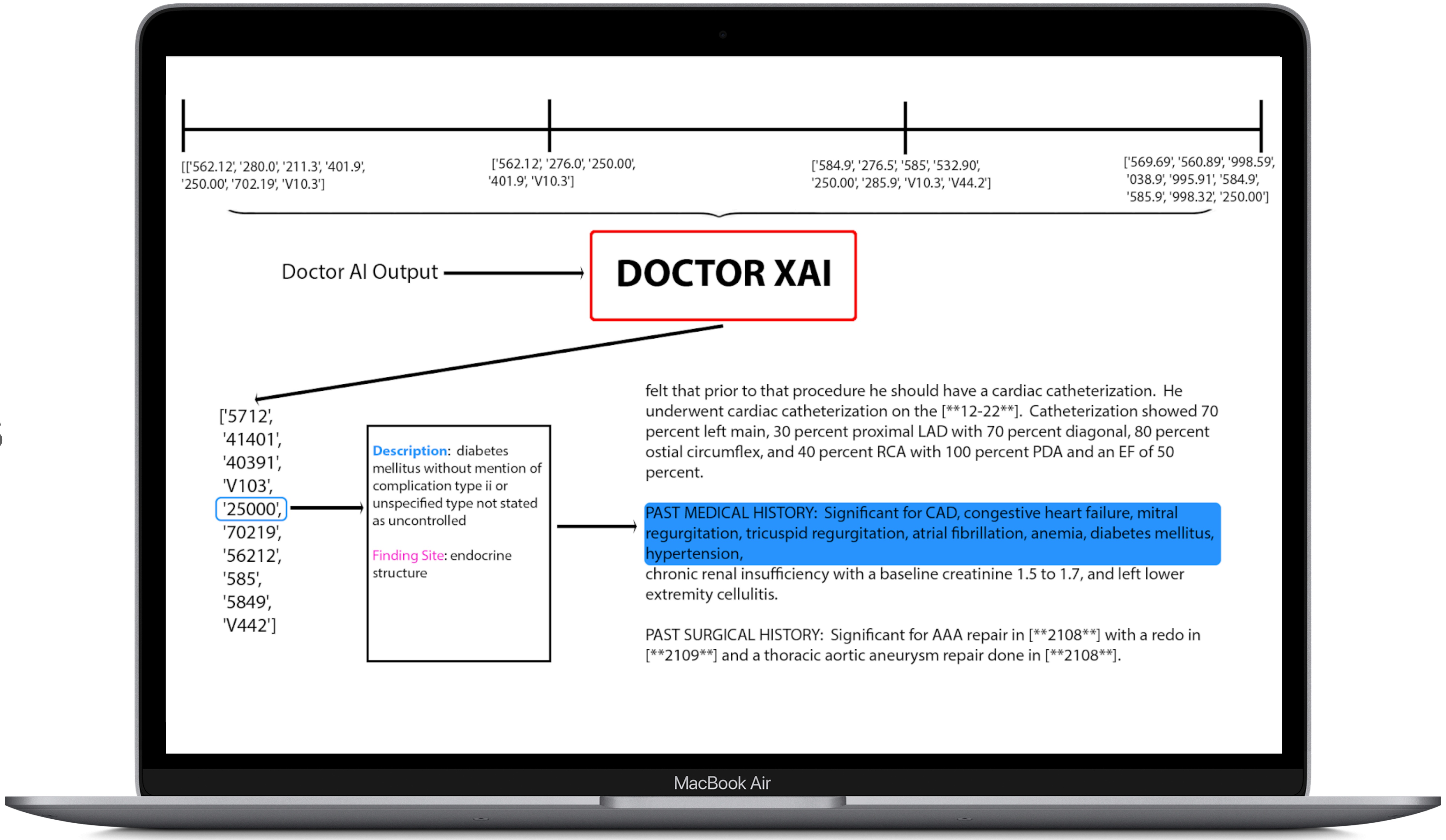
It explains the reason
Behind a prediction

Doctor XAI + Clinical Notes

We used the clinical
Notes to enrich the
Explanation of Doctor XAI

Our goal

Enrich Doctor XAI's
Explanation by highlighting
The most relevant sentences
In the clinical notes



Methodology

We exploit the clinical notes

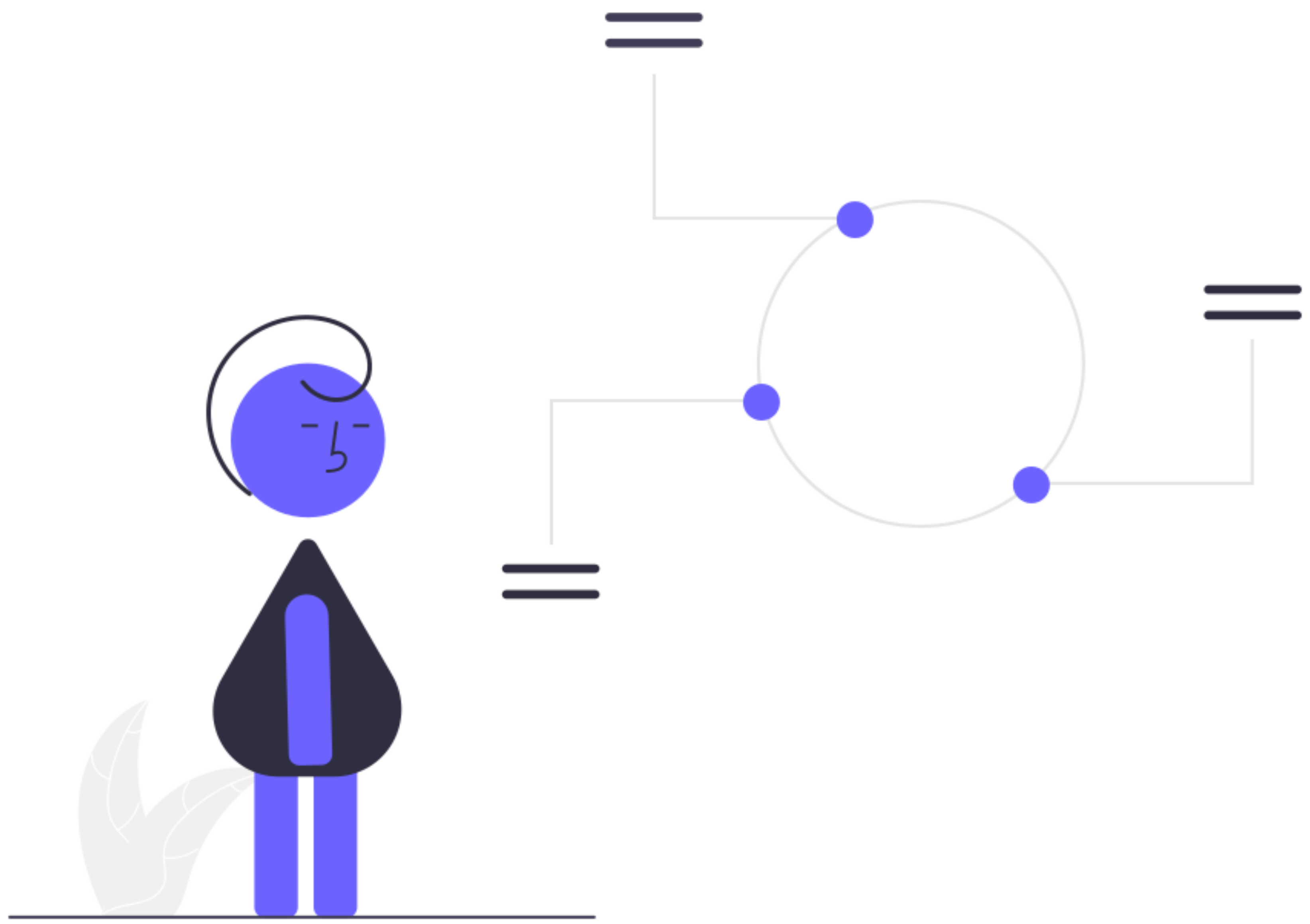
- We used a clinical dataset [1] that contains notes written by clinicians
- A note contains information about patient's clinical history
- We want to extract from the notes the most relevant part for our explanation

[1] MIMIC-III, a freely accessible critical care database - Johnson, Alistair E. W. and Pollard, Tom J. and Shen, Lu and Lehman, Li-wei H. and Feng, Mengling and Ghassemi, Mohammad and Moody, Benjamin and Szolovits, Peter and Anthony Celi, Leo and Mark, Roger G.

How do we extract a sentence from a note?

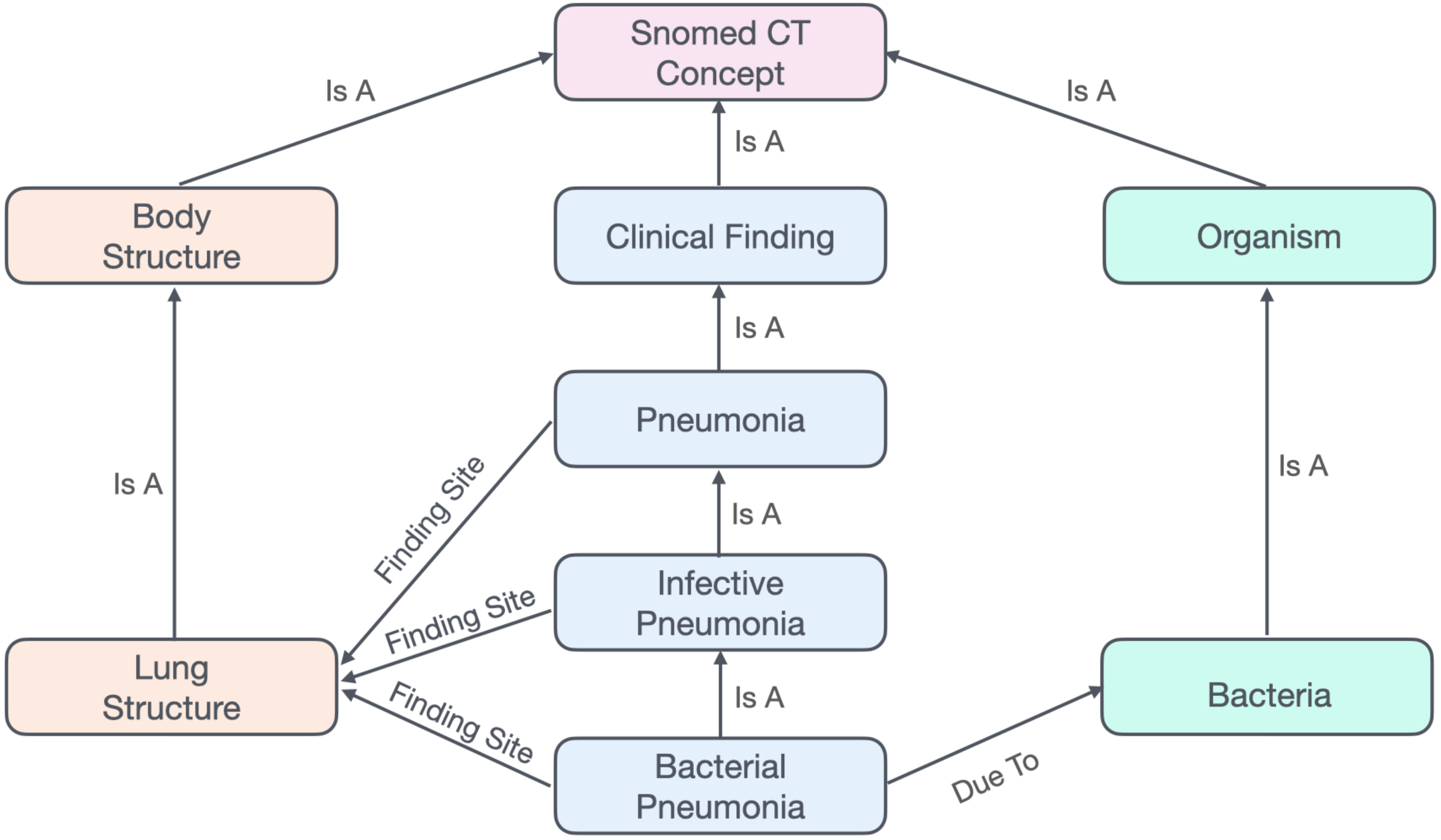
We used the SNOMED-CT
Medical ontology

For each note we highlighted the
Most similar parts to the
Relations taken from the
ontology



The goal is to highlight the
Description, the reason,
The finding site and the
Associated morphology
Of each disease

Snomed-CT



Snomed CT Concept
Is A
Is A
Is A
Body Structure
Clinical Finding
Organism
Is A
Pneumonia
Is A
Finding Site
Is A
Is A
Infective Pneumonia
Finding Site
Lung Structure
Finding Site
Is A
Bacteria
Bacterial Pneumonia
Due To

An example:

Admission Date: [**2163-9-21**] Discharge Date: [**2163-9-27**]

Date of Birth: [**2104-7-1**] Sex: M

Service: MEDICINE

Allergies:
Ceftriaxone

Attending:[**First Name3 (LF) 943**]
Chief Complaint:
fever

Major Surgical or Invasive Procedure:
multiple paracentesis

History of Present Illness:
Pt is a 59 yo man w/ h/o Hep C cirrhosis s/p Liver xplant in [**11-8**], w/ chronic rejection (demonstrated on biopsy in [**9-9**]), recurrent Hep C on INF and ribavirin, B cell lymphoma, who p/w fevers, abdominal pain, SBP. Pt was in USOH until 1 week PTA when began feeling fatigued, had N/V approximately 1-2 episodes per day, non-bloody, non-bilious. 3 days PTA, pt began to have severe abdominal pain. He also noted increased abd girth, increased LE edema, R > L, denied any calf pain. Over past 3 days, pt also c/o cough with some sputum production, although difficult to bring up 2/2 abd pain. He also c/o laryngitis starting 3 days ago. ROS otherwise negative for BRBPR, melena, SOB, CP/pressure."

Description: “Anemia in chronic kidney disease”

[0.314, -1.456, ….., 3.5644, 7.54432]

“Chronic Kidney Disease Stage II”

[0.436, 7.655, ….., -4.2533, 1.78824]

Compute the distance between the two embeddings

What embeddings we used?

BioWordVec

A pre-trained word2vec word embedding for biomedical natural language processing trained Mimic-III

BioSentVec

A biomedical sentence Embedding with sent2vec Trained on Mimic-III

Clinical Bert

A Bert based embedding Trained on Mimic-III

Results

How we validated our model

- We did not find any annotated clinical dataset suitable for our task
- A domain expert annotated 32 clinical notes by highlighting the relevant sentences
- We compared the manually annotated notes with the sentences extracted with our method

Model Validation

- BioWordVec is (surprisingly) the best word embedding
- The “Description” relation is the easiest to highlight
- It is not easy to deal with “Finding Site” and “Due To”

Relationship	Embedding		Accuracy	F1-score	Precision	Recall
Description	BioWordVec	Value	**0.718**	**0.707**	**0.819**	**0.622**
		Confidence	0.715 - 0.719	0.704 - 0.708	0.815 - 0.819	0.619 - 0.624
Description	BioSentVec	Value	0.662	0.664	0.804	0.566
		Confidence	0.659 - 0.663	0.661 - 0.665	0.800 - 0.804	0.563 - 0.567
Description	ClinicalBert	Value	0.640	0.602	0.654	0.557
		Confidence	0.637 - 0.641	0.599 - 0.603	0.651 - 0.655	0.555 - 0.559
Finding site	BioWordVec	Value	**0.743**	0.274	0.170	**0.708**
		Confidence	0.740 - 0.744	0.273 - 0.277	0.169 - 0.173	0.705 - 0.709
Finding site	BioSentVec	Value	0.726	**0.294**	**0.200**	0.555
		Confidence	0.723 - 0.727	0.293 - 0.297	0.199 - 0.203	0.553 - 0.557
Finding site	ClinicalBert	Value	0.686	0.214	0.150	0.375
		Confidence	0.683 - 0.687	0.213 - 0.217	0.149 - 0.153	0.373 - 0.377
Due to	BioWordVec	Value	**0.666**	**0.451**	**0.350**	**0.636**
		Confidence	0.647 - 0.673	0.440 - 0.466	0.342 - 0.368	0.618 - 0.644
Due to	BioSentVec	Value	0.600	0.091	0.050	0.500
		Confidence	0.582-0.609	0.091 - 0.119	0.050 - 0.080	0.486 - 0.513
Due to	ClinicalBert	Value	0.568	0.214	0.150	0.375
		Confidence	0.552 - 0.579	0.211 - 0.238	0.149 - 0.176	0.366 - 0.392
Associated morphology	BioWordVec	Value	**0.856**	**0.577**	**0.464**	**0.764**
		Confidence	0.845 - 0.856	0.571 - 0.581	0.459 - 0.470	0.755 - 0.766
Associated morphology	BioSentVec	Value	0.803	0.409	0.321	0.562
		Confidence	0.793-0.803	0.405 - 0.415	0.318 - 0.329	0.556 - 0.566
Associated morphology	ClinicalBert	Value	0.734	0.339	0.321	0.360
		Confidence	0.726 - 0.736	0.336 - 0.347	0.318 - 0.329	0.356 - 0.367

Table 1. Validation of our methodology on 32 manually annotated clinical notes of 9 patients. Confidence of Accuracy, Precision, Recall and F1-score at $1 - \alpha = 0.95$ of confidence level.

Conclusions

Conclusions

- We presented a method to semantically enrich a XAI explanation in the healthcare context

- We performed some experiments annotating a part of a popular dataset

- We studied several approaches to extract the information from the notes and we compared different embeddings

Future works

- Validate the methodology on a larger quantity of clinical notes
- Test the methodology to understand if the semantically enriched explanation improves the interpretability of Doctor AI
- We would like to investigate the opportunity to exploit our methodology to generate explanation expressed by natural language

Thank you!

📧 luca.corbucci@phd.unipi.it