Automated Feature Extraction for the STS National Database: The Impact of Artificial Intelligence

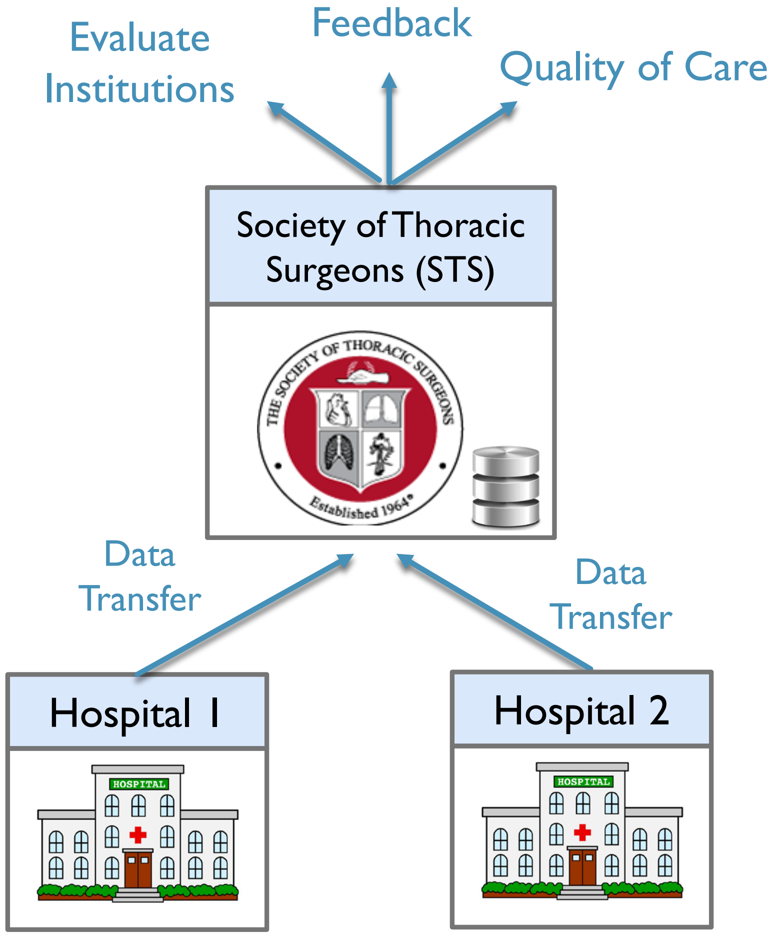
A joint effort between:
Giorgos Margaritis*, Pericles Petrides*, Dimitris Bertsimas, Robert Habib,
Agni Orfanoudaki, David Shahian

October 16th 2023
Informs 2023

THE CHALLENGES OF IMPROVING THE STS DATABASE

- US Hospitals submit data to National Registries for:
  - Hospital evaluation
  - Quality of care

- Society of Thoracic Surgeons (STS):
  - Gold-standard national database
  - Penetration in almost all cardiac programs (97%)
  - Goals:
    - Evaluate & Compare institutions/programs
    - Provide feedback
    - Improve Quality of Care



More than 1000 variables!

An STS Record

MRN	Age	Diabetes	Hypertn	ChrLungD	…
12345	25	1	1	0	…
		…			

A Deep-Dive into the Current Workflow of Data Managers

Data Manager

Cardiac Surgery Patient

Epic

History & Physical (H&P)

Time: Oct 12th at 11:48
John Doe, 32 y.o. male
History of Present:
Long history of diabetes

Operative Note (Opn)

Time: Oct 13th at 17:00
The patient underwent coronary bypass surgery under general anesthesia.

• • •

Lab tests (Labs)

Time: Oct 13th at 09:05
- HbA1c : 5.7
- White blood cells: 6k
- Platelet count: 200k

Medication (Med)

Time: Oct 14th at 11:32
- Insulin injection 100mg
- Lisinopril 20mg
- Augmentin 300mg

Unstructured (text) Records

Structured Records

THE CHALLENGES OF IMPROVING THE STS DATABASE

- Translating EHR Reports to STS format:
  - **Manual work** from hospitals & STS
  - High operational cost for collaborating hospitals
  - Requires trained and qualified personnel
- Overall, a significant **operational overhead.**

STS needs more variables to improve quality measures and risk scores

Participants want to reduce the data collection burden.

Can we create a generalizable and standardized process using all the available sources for as many variables as possible?

OBJECTIVE: EXTRACT INFORMATION FOR MULTIPLE VARIABLES FROM MULTIPLE SOURCES

EPIC Data Sources
(Structured and Unstructured Text Files)

	Data Source
unstructured	Operative Notes
	History & Physical Reports
	Cardiology Reports
	Pulmonary Notes
	Pathology Notes
	Radiology Reports
	Endoscopy Notes
structured	Diagnoses
	Medicines
	Labs
	…

- ***10+** different sources*
- ***9,000** patients, multiple visits*
- *~**20GB** of data*

STS Database
(Tabular data with **1000+** variables of interest)

Outcome	
Diabetes	Pre-operative
Hypertension	
Chronic Lung Disease	
Peripheral Arterial Disease	
CABG Operation	Intraoperative
Aortic Valve Procedure Atrial Fibrillation	
Post-operative stroke	Post-operative
Post-operative unplanned Aortic Intervention	
Expired in Operating Room	
…	

WHAT IF WE ONLY HAVE 1 SOURCE?

PREDICT ONE VARIABLE USING A SINGLE SOURCE

Patient #1

History & Physical (H&P)

Patient Information:
John Doe, 32 y.o. male

History of Present Illness:
Long history of diabetes

Patient #2

History & Physical (H&P)

Patient Information:
Jack Miller, 23 y.o. male

History of Present Illness:
No presence of diabetes

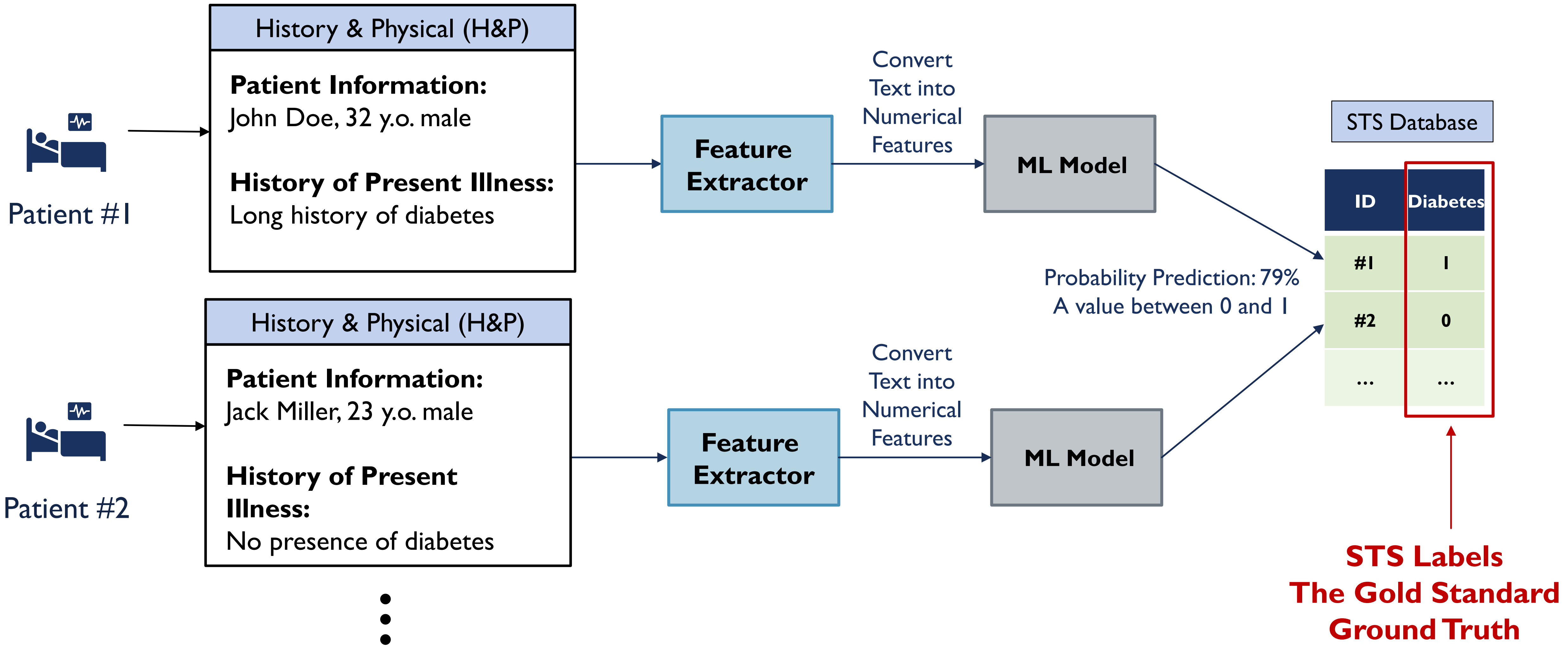
⋮

STS Database

ID	Diabetes
#1	1
#2	0
…	…

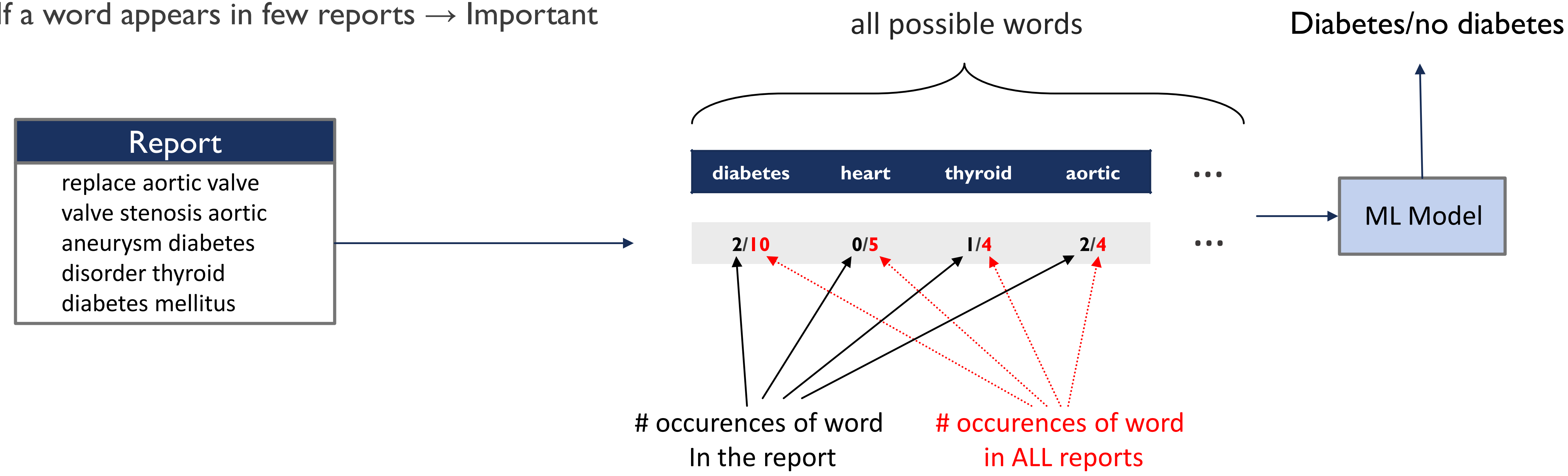
STS Labels
The Gold Standard
Ground Truth

PREDICT ONE VARIABLE USING A SINGLE SOURCE



FEATURE EXTRACTOR 1: TF-IDF

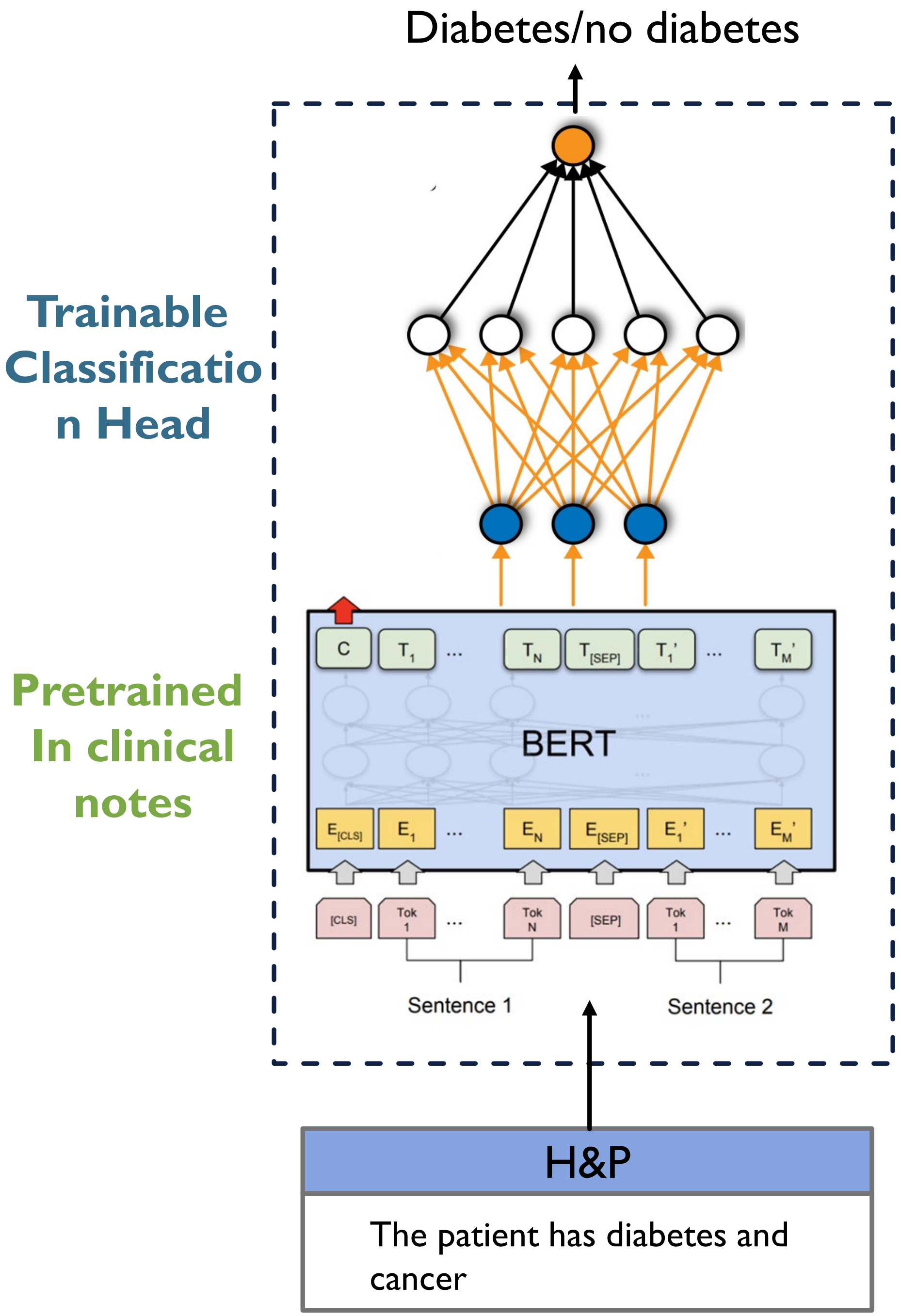
- Simple technique and very fast
- Does not consider word order but only word frequency
- If a word appears in many reports → Not very important
- If a word appears in few reports → Important



FEATURE EXTRACTOR 2: CLINICAL BERT

- ClinicalBERT:
  - Transformer-based Large Language Model
  - Trained on MIMIC-III to understand medical text
  - Accounts for **context** → Contextualized embeddings

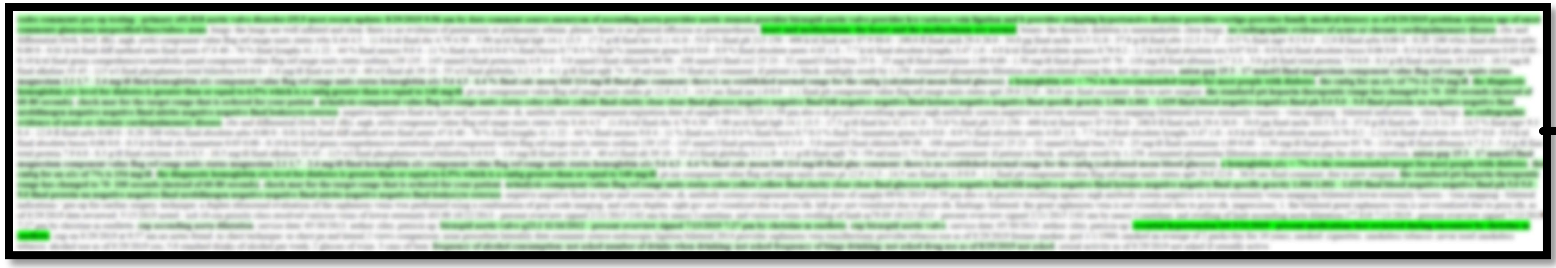
- Predict with ClinicalBERT → **Fine-tuning:**
  - Take a pre-trained BERT model
  - Add a classification NN on top (CLS head)
  - Train BERT+CLS head for few epochs on our data



S-BERT: USING AI TO SUMMARIZE LARGE REPORTS

- Some reports are very big and cannot be used as an input to the ClinicalBERT model (max input 500 words).
- For those reports, we apply a different AI model to summarize their content and keep only relevant sentences to the variable of interest.

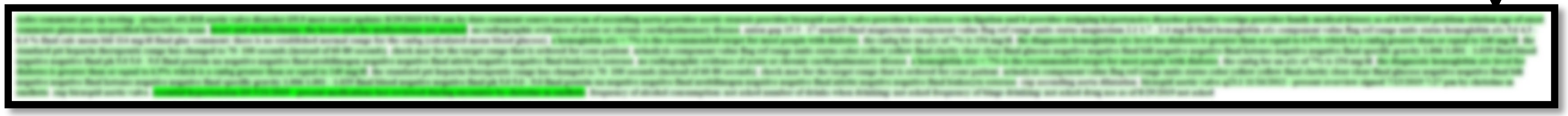
Step 1: Find similar sentences to the Target Sentence (STS Manual): *"Diabetes, mellitus, blood glucose, hemoglobin A1c, HbA1c"*



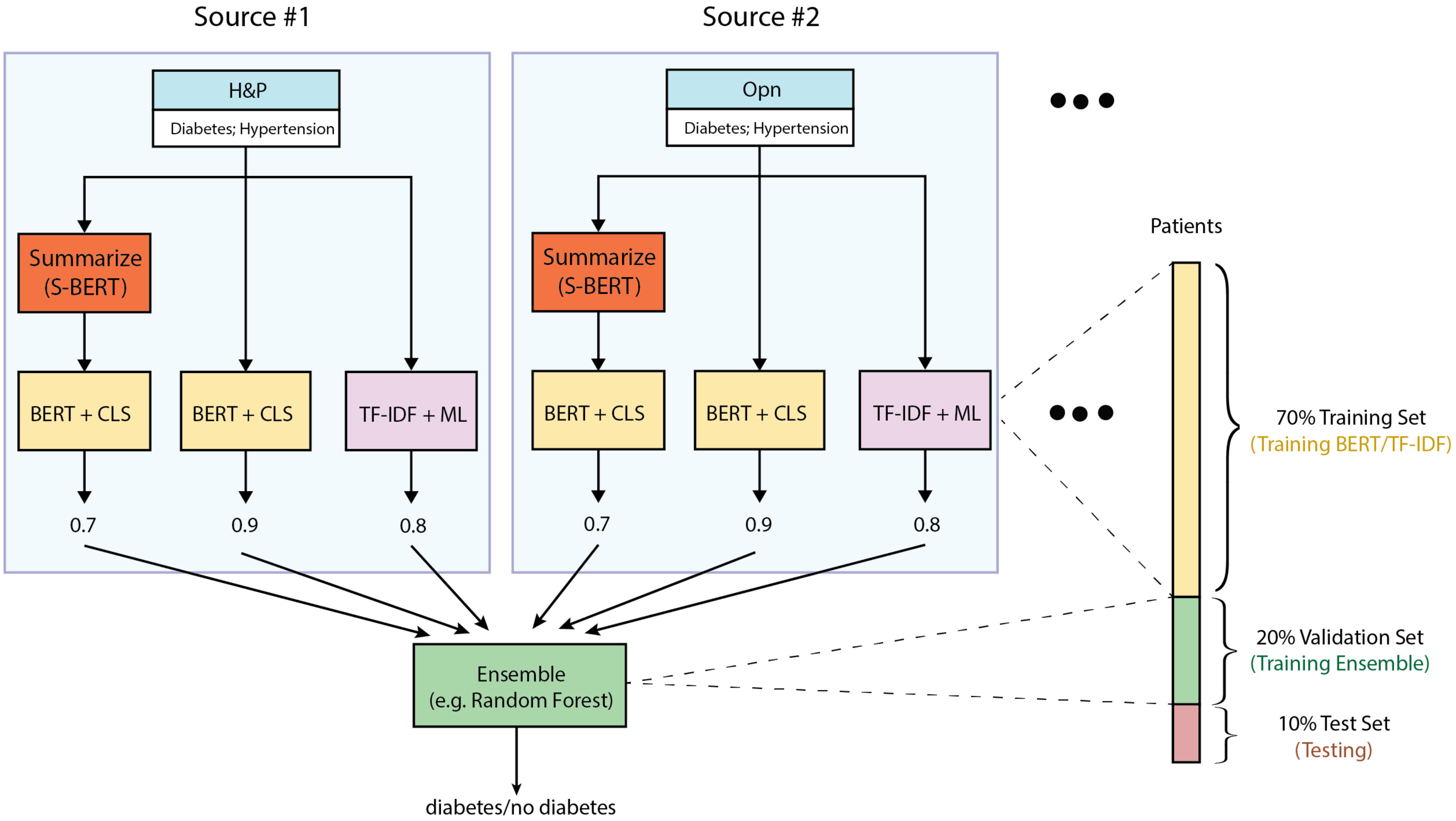
Visit Note (~1700 words)

S-BERT

Step 2: Keep "most similar" sentences (until 500 words are filled)



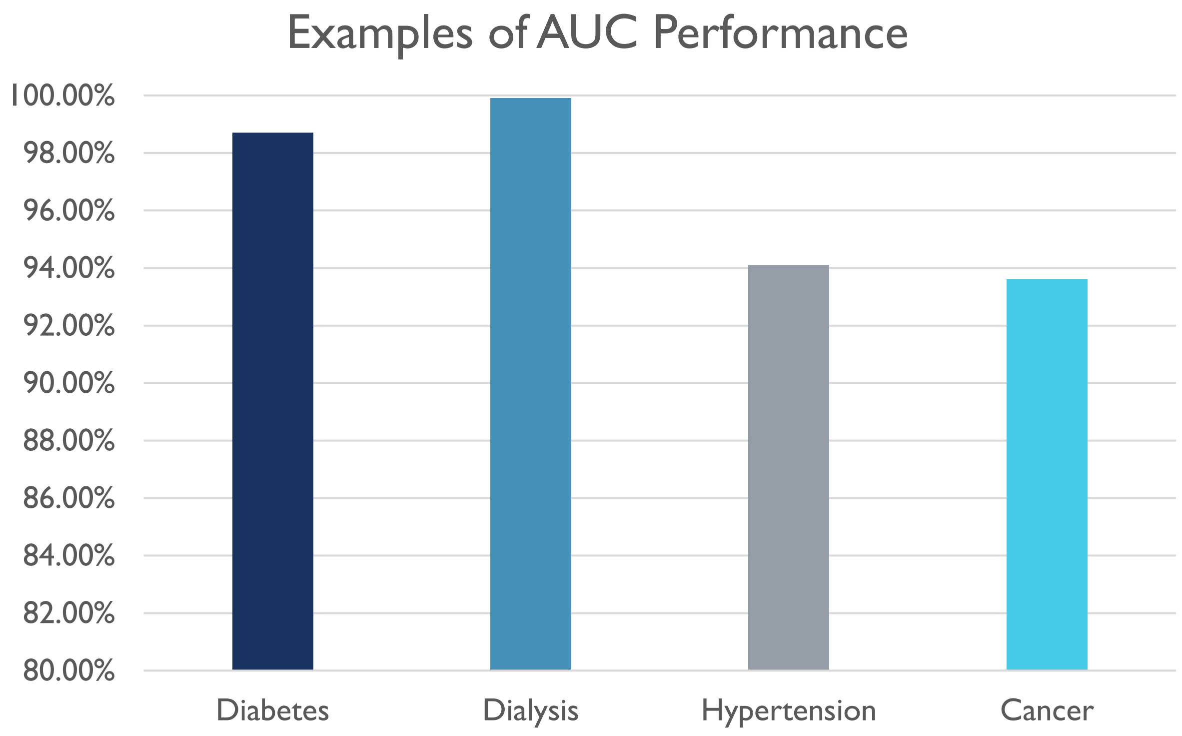
"Summary" (500 words)



Source #1
Source #2
H&P
Diabetes; Hypertension
Opn
Diabetes; Hypertension
Summarize (S-BERT)
BERT + CLS
BERT + CLS
TF-IDF + ML
Summarize (S-BERT)
BERT + CLS
BERT + CLS
TF-IDF + ML
0.7
0.9
0.8
0.7
0.9
0.8
Ensemble (e.g. Random Forest)
diabetes/no diabetes
Patients
70% Training Set
(Training BERT/TF-IDF)
20% Validation Set
(Training Ensemble)
10% Test Set
(Testing)

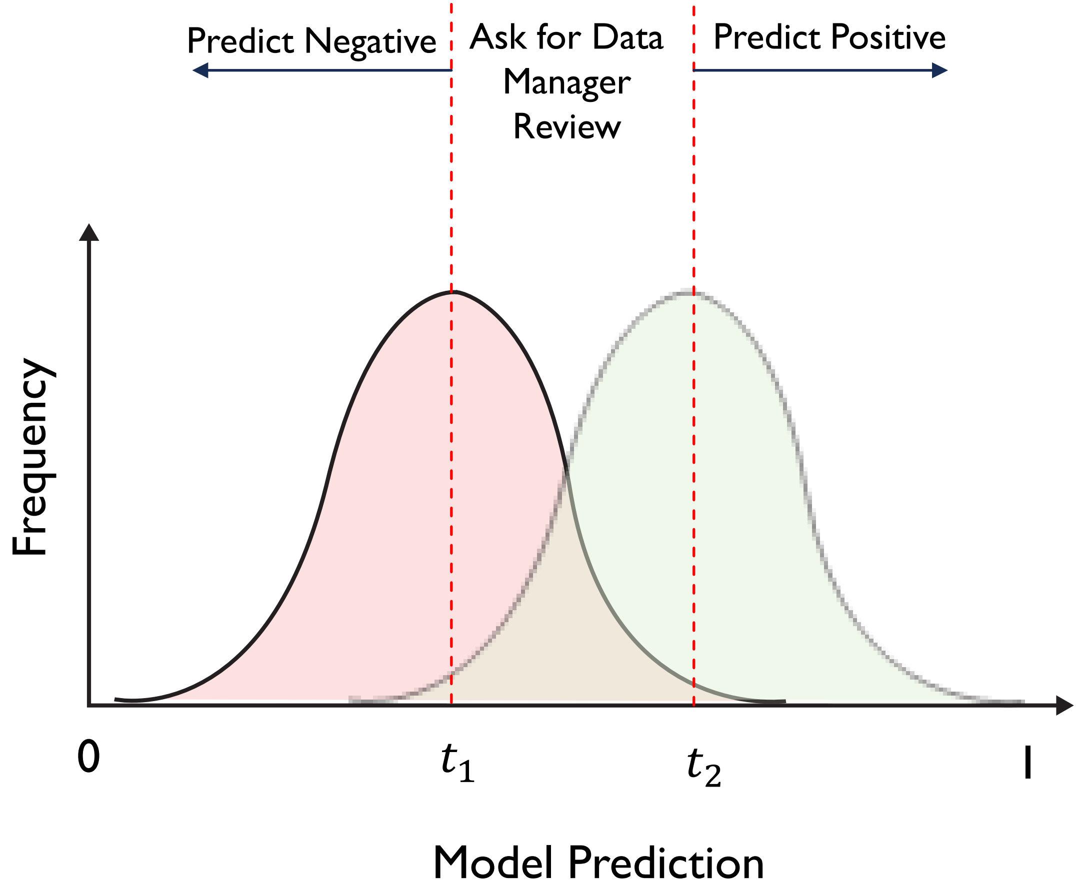
ACCURACY-COMPLETION TRADE-OFF

- Measure performance with **AUC:**
  - Also accounts for class imbalance/sparse variables
- Different performance on each outcome
- Issue:
  - Very stringent performance requirements set by doctors (i.e. more **95-97%** AUC)
  - What happens if we don't meet them for a particular outcome?

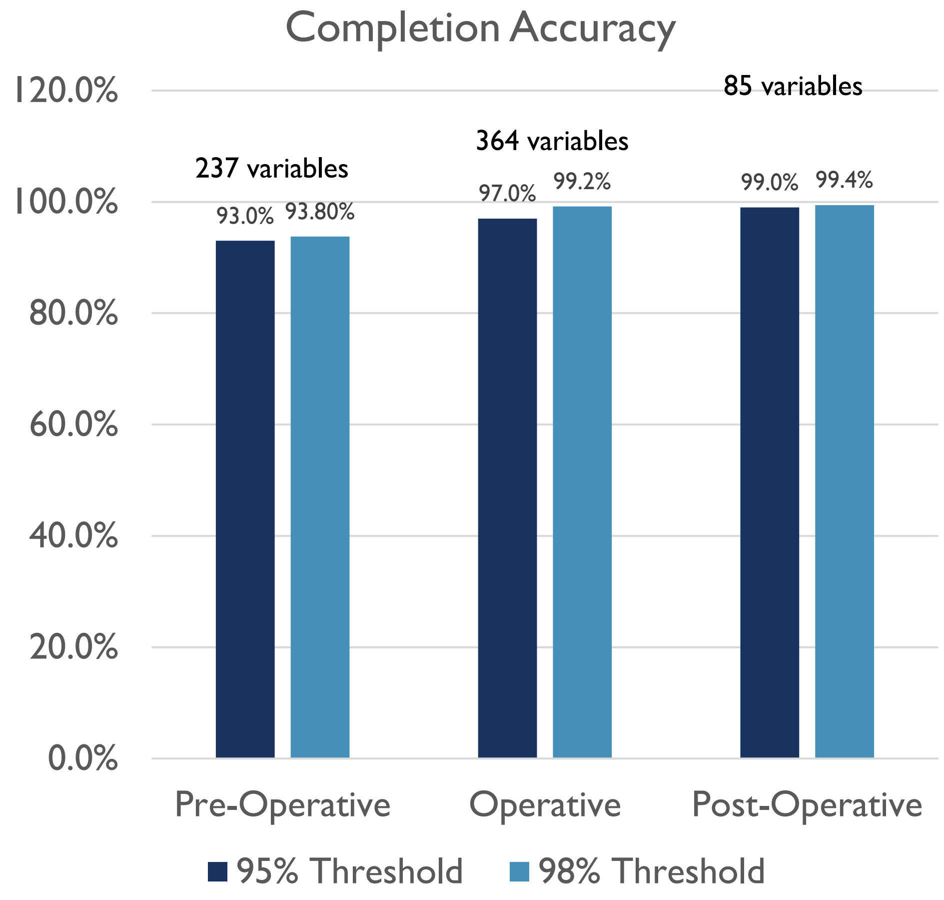


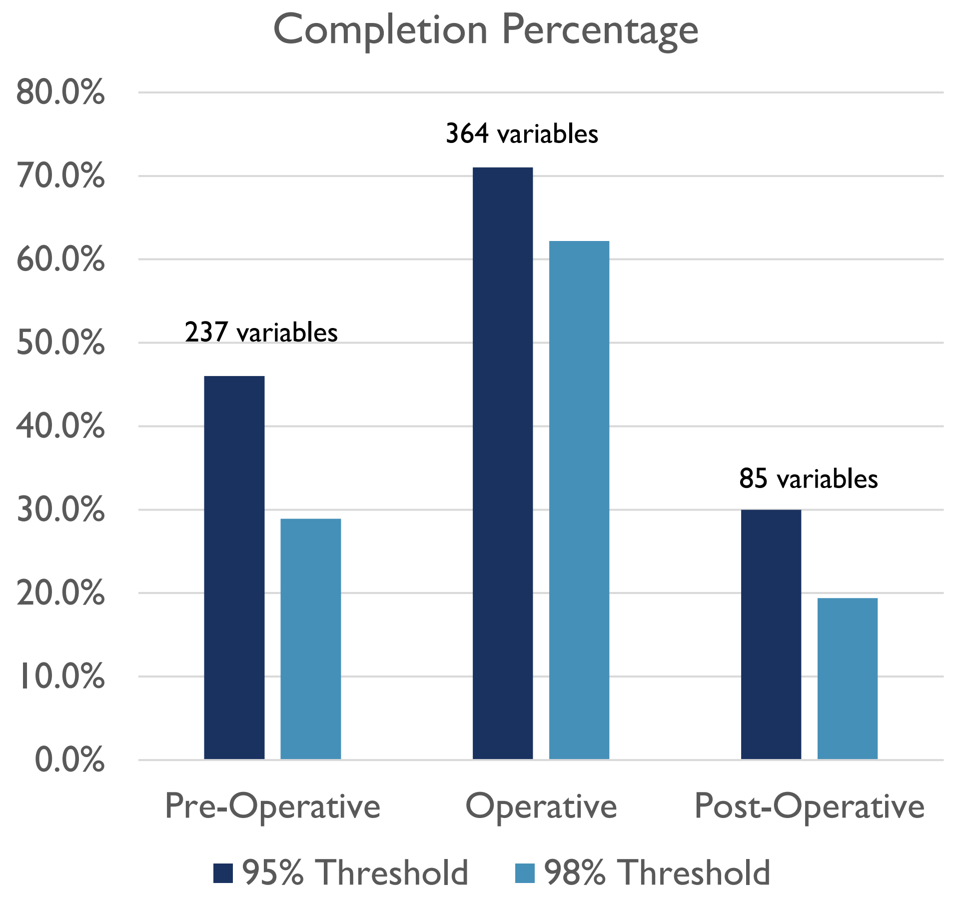
PREDICTING ONLY WHEN WE ARE CONFIDENT

- Find $\boldsymbol{t_1}$ such that we make few mistakes when classifying **negative** below $\boldsymbol{t_1}$
- Find $\boldsymbol{t_2}$ such that we make few mistakes when classifying **positive** above $\boldsymbol{t_2}$
- If between $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$, leave prediction to the Data Manager
- Higher accuracy requirement:
  - $\boldsymbol{t_1}$ closer to 0 and $\boldsymbol{t_2}$ closer to 1
  - More predictions left to the Data Manager
  - Accuracy/Completion trade-off



OVERALL PERFORMANCE PIPELINE: ACCURACY VS COMPLETION RATE THE VARIABLE VIEW

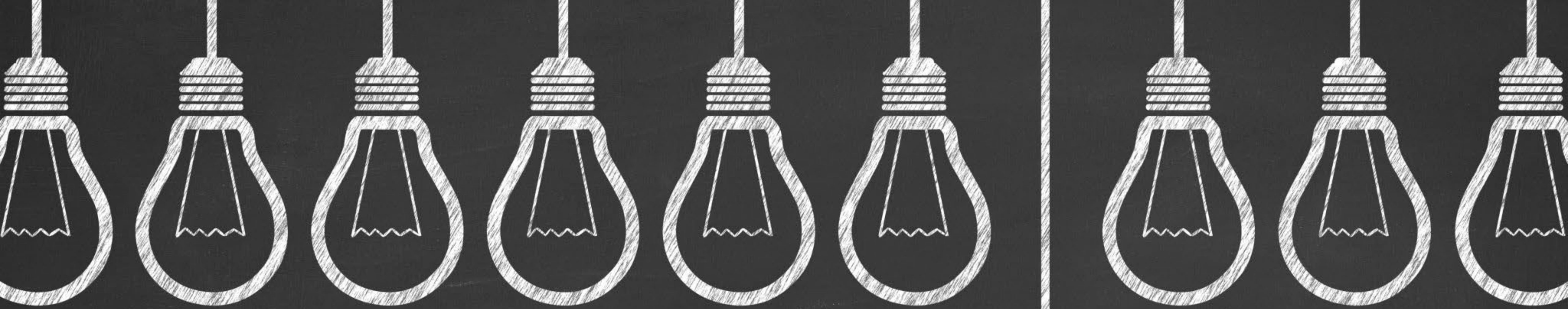




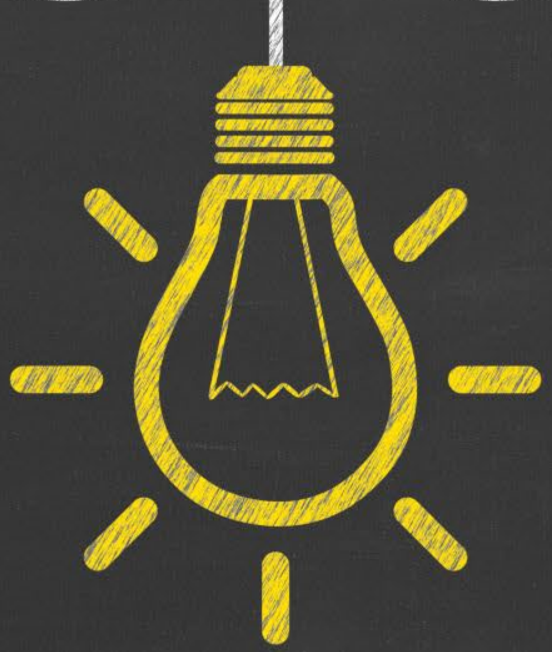
Overall Accuracy: **97.6%** **98.1%**

AI Models AUC: **95.2%** **96.5%**

Overall Completion %: **47.2%** **43.3%**



What if we exclude AI-predicted variables with final completion accuracy less than the STS standard of 96%?



- ***Overall Completion Rate***: *40.8%*
- ***Completion Accuracy***: *99.7%*

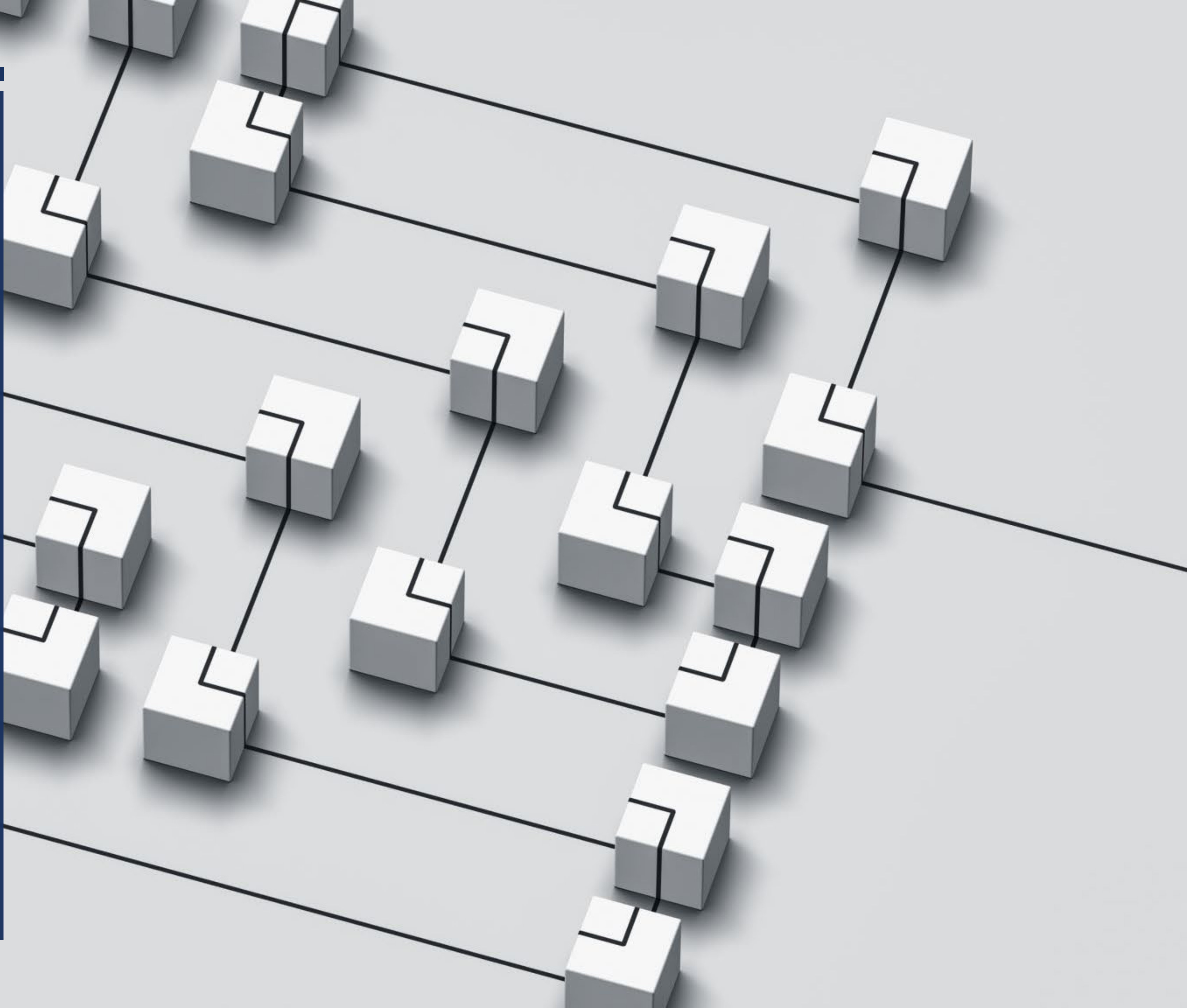
CONCLUSIONS

Our proposed AI-based pipeline can:

- Lead to substantial reductions of the data collection burden (at least 40%)
- Maintain the high-standards of quality and accuracy (at least 99%)
- Offer high levels of automation without human involvement (at least 40%)
- Improve discrepancies and standardize the database input
- Provide a paradigm for other national registries

NEXT STEPS

- Extension of the AI pipeline to a wider set of variables
- External validation of the pipeline to Hartford Healthcare



Thank you!!

Questions?