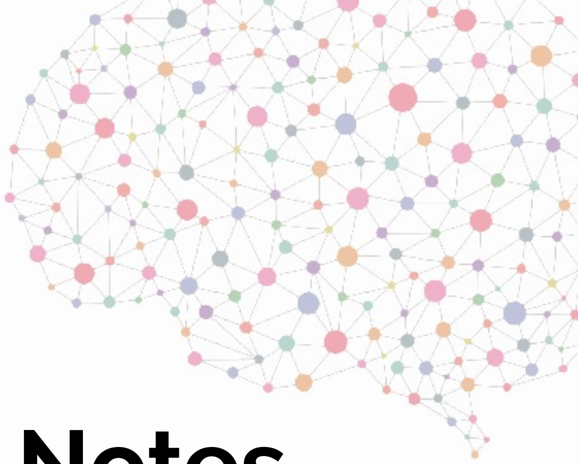
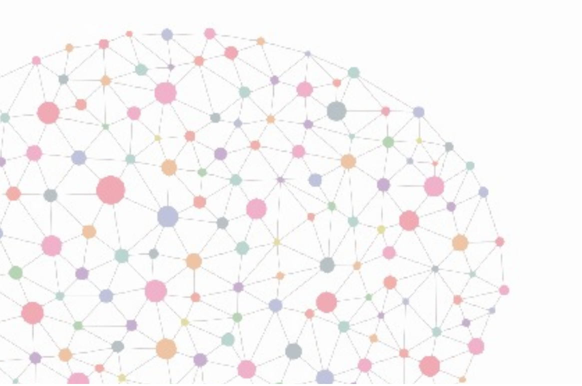


Predicting Psychiatric Readmission from Clinical Notes

Emmy Phung & Kevin Yie

Overview



- Introduction
- Purpose & Hypothesis
- Data
- Materials & Methods
  - Models
  - Training
- Results
- Discussion
  - Bert-viz
  - Future Direction

Introduction

- Attempts on predicting heart failure readmission from clinical notes have been successful
  - Model: Random Forest, 3-layer CNN
  - Embedding method: word2vec trained on PubMed abstracts and PubMed Central full text articles
- Attempts to predict psychiatric readmission show great potential
  - Mostly traditional methods such as LDA
  - Due to how recent it was, few published literature using ELMo and BERT or Transformers
- Fine-tuning BERT on clinical discharge notes have shown success (Bio Discharge BERT)
  - NER, De-identification, and natural language inference tasks

Hypothesis

We believe that the stages of a **mental illness** cannot be fully described through numerical measurements and are **best captured in words.**

Deep learning models will allow us to **predict psychiatric readmission** and **extract meaningful insights** from **clinical notes .**

Purpose

Rapid psychiatric readmission have been associated with lack of **family support**, **severity of symptoms**, and **premature discharge.**

Due to the many factors that impact readmission rate, we want to create a model that assists clinicians in determining if a patient is at **risk of rapid readmission upon discharge.**

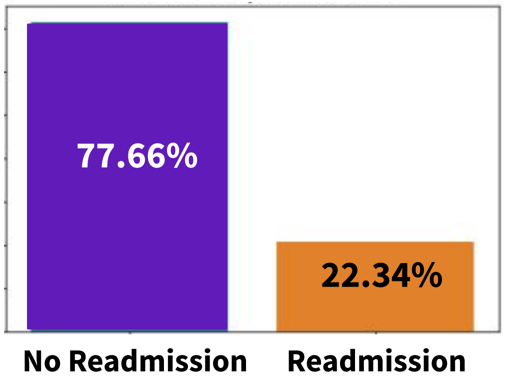
Data - MIMIC III

- Two Tables:
    - Diagnosis ICD (per visit)
        - 290 - 319 (mental disorders) and E9500-E9600 (self-harm)
    - Note Events (per visit)
        - Text data extracted using regex patterns

- SQL filtering and joining to form final database with granularity at admission level
- Labelling (per admission):
    - **Readmission : 1 (22.34%)** *(if the admission is **followed by another admission**)*
    - **No readmission : 0 (77.66%)** *(if the admission is **not followed by another admission**)*

All Diseases	Mental Disorders
Number of unique patients	
46,520	7,050
Number of admissions	
58,976	7,958
Number of discharge notes	
59,652	9,218

Class size

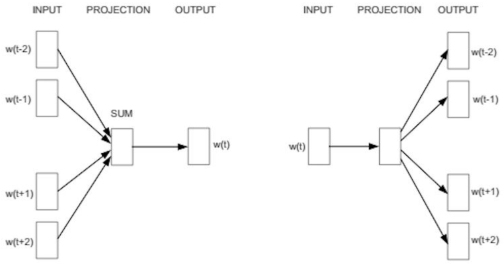


Our Models

- **LSTM** with no pre-trained embedding
- **LSTM** with GloVe embedding
- **BERT-base**
- **Bio Discharge BERT**

Vector Embeddings

- **Word2Vec**
  - Allows us to do vector operations e.g. King - man + woman = Queen
  - Limitations: only considers local context



CBOW Skip-gram

- **GloVe**
  - Uses global co-occurrence to create embeddings out of conditional probability task
  - Seek to predict term P_ik/P_jk
  - Limitations: 1 vector per word
    - E.g. “jail cell” vs “animal cell”

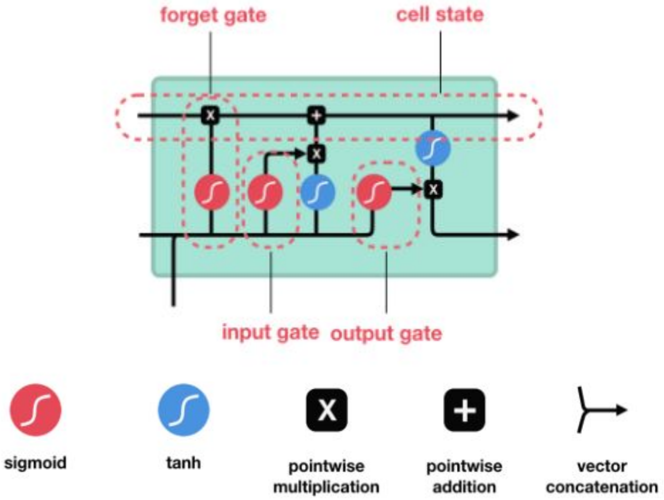
	the	cat	sat	on	mat
the	0	1	0	1	1
cat	1	0	1	0	0
sat	0	1	0	1	0
on	1	0	1	0	0
mat	1	0	0	0	0

Probability and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36	0.96

https://towardsdatascience.com/light-on-math-ml-intuitive-guide-to-understanding-glove-embeddings-b13b4f19c010

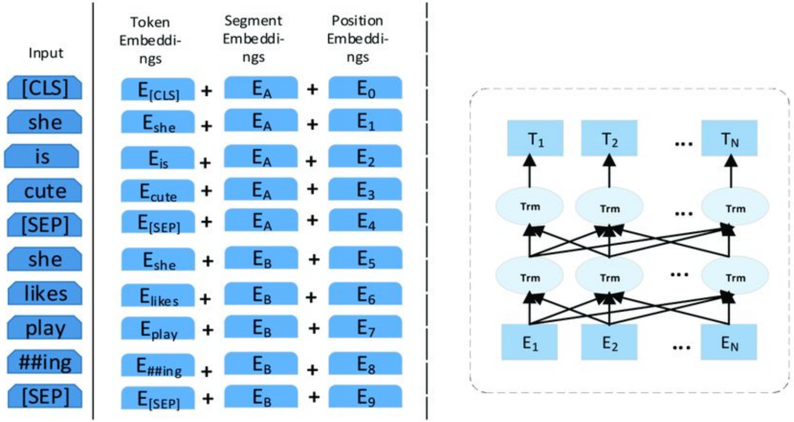
Models

LSTM



- Very natural application for text sequences
- "Forgets" and "Updates" information at each token
- Limitations: "Black Box" and depends solely on cell state/output from previous cell

BERT



- Context considered through positional embeddings
  - Results in possible different vectors for same word
- Attention considers all tokens rather than just previous layer's information

Training

https://ruder.io/state-of-transfer-learning-in-nlp/

- **LSTM w/ FC Linear Classification Layer**
  - 2 configs: **No pretrained embedding** vs. **GloVe embedding**
  - WordPiece tokenizer
  - LSTM hidden dimension : 40, 60, 80, 100

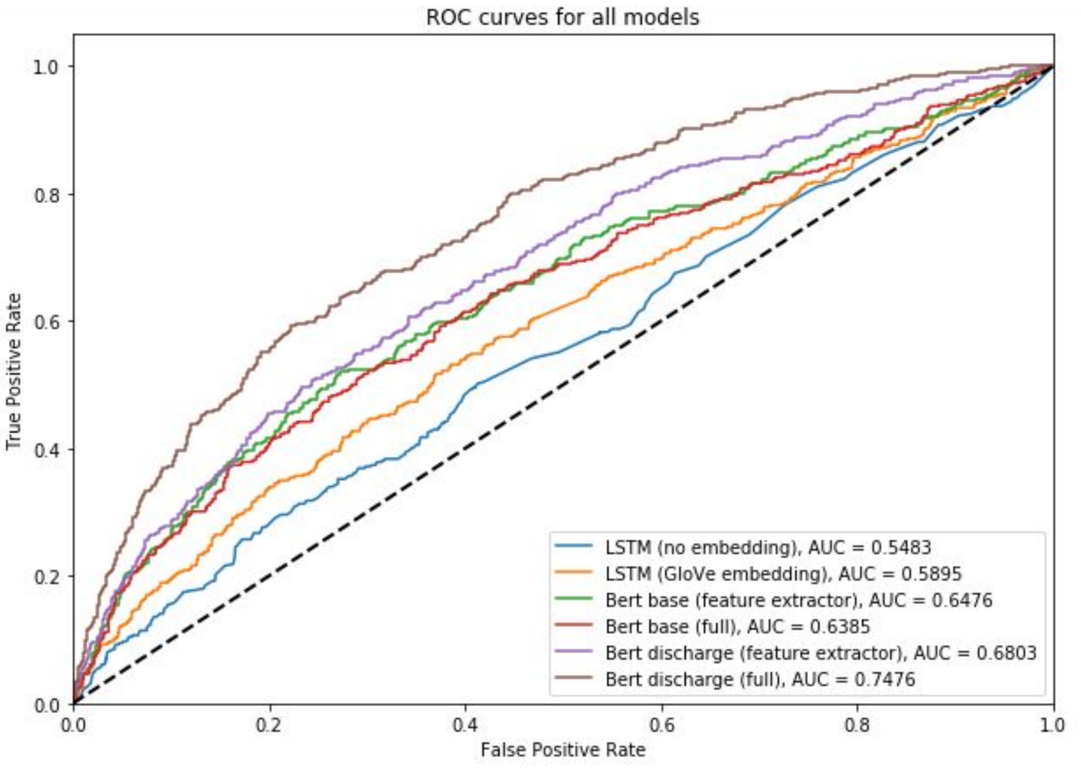
- **BERT base (pretrained)** vs. **Discharge BERT (pretrained on Discharge Summary)**
  - Bert Tokenizer (bert-base-uncased)
  - AdamW opitimizer
  - Use BERT as **feature extractor (freeze Bert layer)**
    - Learning rate: 2e-2, 2e-3, 2e-4, 2e-5
    - Batch size: 12, 32
  - **Train Bert layer w/ final classifier layer**
    - Learning rate: 2e-5, 3e-5, 4e-5, 5e-5
    - Batch size: 6, 12

Evaluation

- Train: **BCE & BCEwithLogits loss**, **AUC**
- Val/test: **Accuracy, AUC, Precision & Recall, F1**

Results

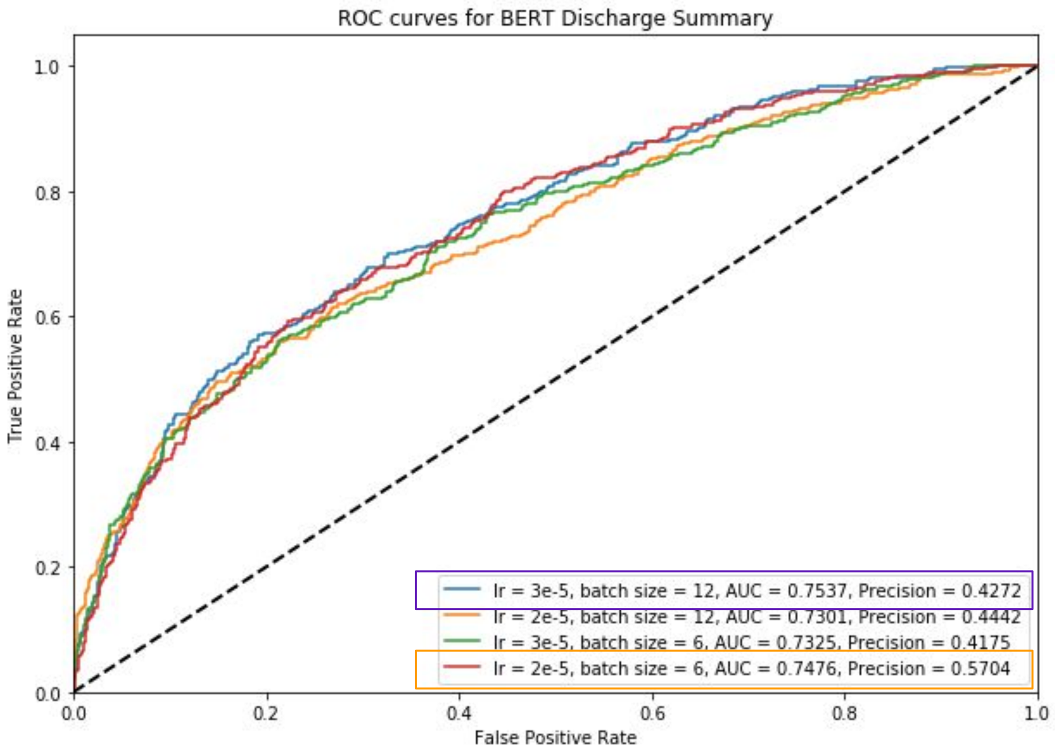
Compare all models by AUC



- Based on AUC, **finetuned BERT Discharge** (pretrained on Discharge Summary notes from MIMIC III) performs the **best (AUC = 0.75)**
- BERT base has more predictive power than LSTM

Result

Hyper parameter tuning



Primary hyper parameters of interest:

- **Learning rate (2e-5 -> 5e-5)**
- **Batch size (6 -> 12)**

When training per epoch, the best model is chosen based on AUC

When comparing across different configs, the best is chosen based on both AUC and precision

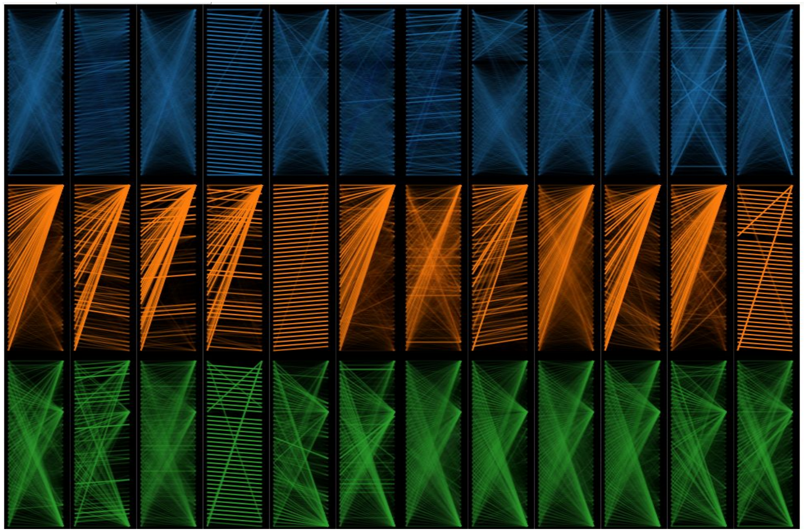
- **Best AUC: 0.75** (lr = 3e-5, batch = 12)
- **Best precision: 0.57** (lr = 2e-5, batch = 6)
- Final model is picked based on precision

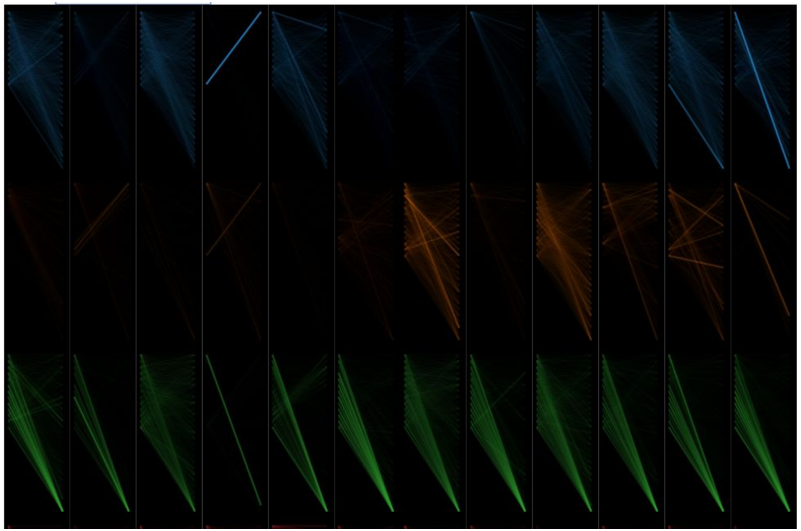
Final results (on test set)

	Accuracy	AUC	Precision	Recall	F1
LSTM (No embedding)	0.7776	0.5483	0.0	1.0	0.0
LSTM (GloVe embedding)	0.7776	0.5895	0.0	1.0	0.0
BERT Base (feature extractor)	0.7798	0.6476	0.1335	0.9630	0.2345
BERT Base (full)	0.7777	0.6385	0.3835	0.9225	0.5418
BERT Discharge Summary (feature extractor)	0.6464	0.6803	0.5850	0.6641	0.6220
BERT Discharge Summary (full)	**0.7402**	**0.7476**	**0.5704**	**0.7891**	**0.6622**

- **BERT Discharge Summary *(full - training all parameters)*** is the **best model** based on on both AUC and F1 score.
- **BERT Discharge Summary *(as feature extractor)*** gives the **best precision** with a sacrifice in accuracy and recall.
- **LSTM** predicts all to be the majority class and **does not have any predictive power**
- In conclusion, the results show that our models have the capacity to learn but there is room for further improvements

Attention and Bert-Viz





- 12 layers with 12 heads each (only 3 heads per layer shown)
- Notice the different patterns we see forming

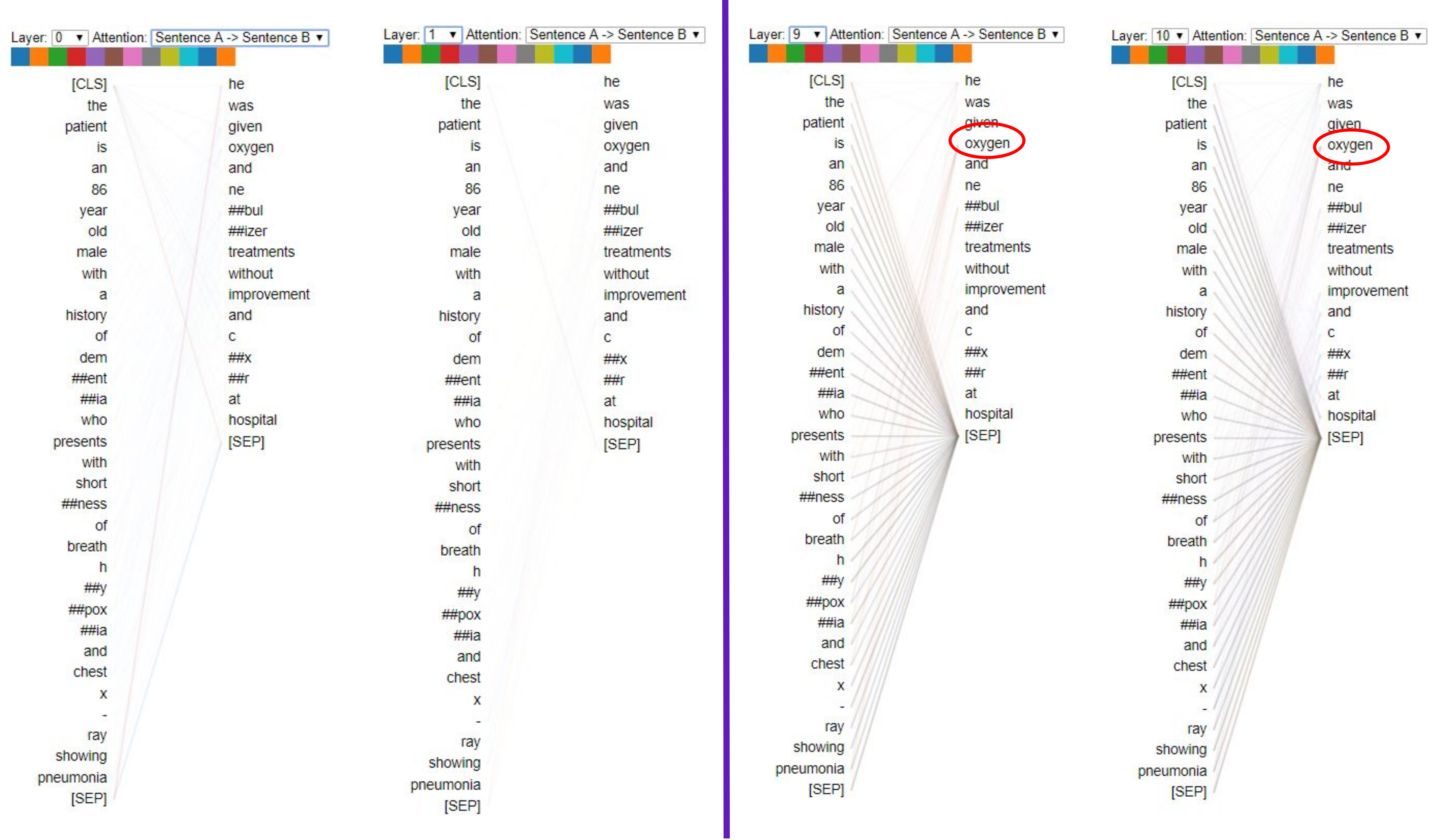
- Attention of sentence A to sentence B (uni-directional)
- Notice that the later layers have more interesting patterns

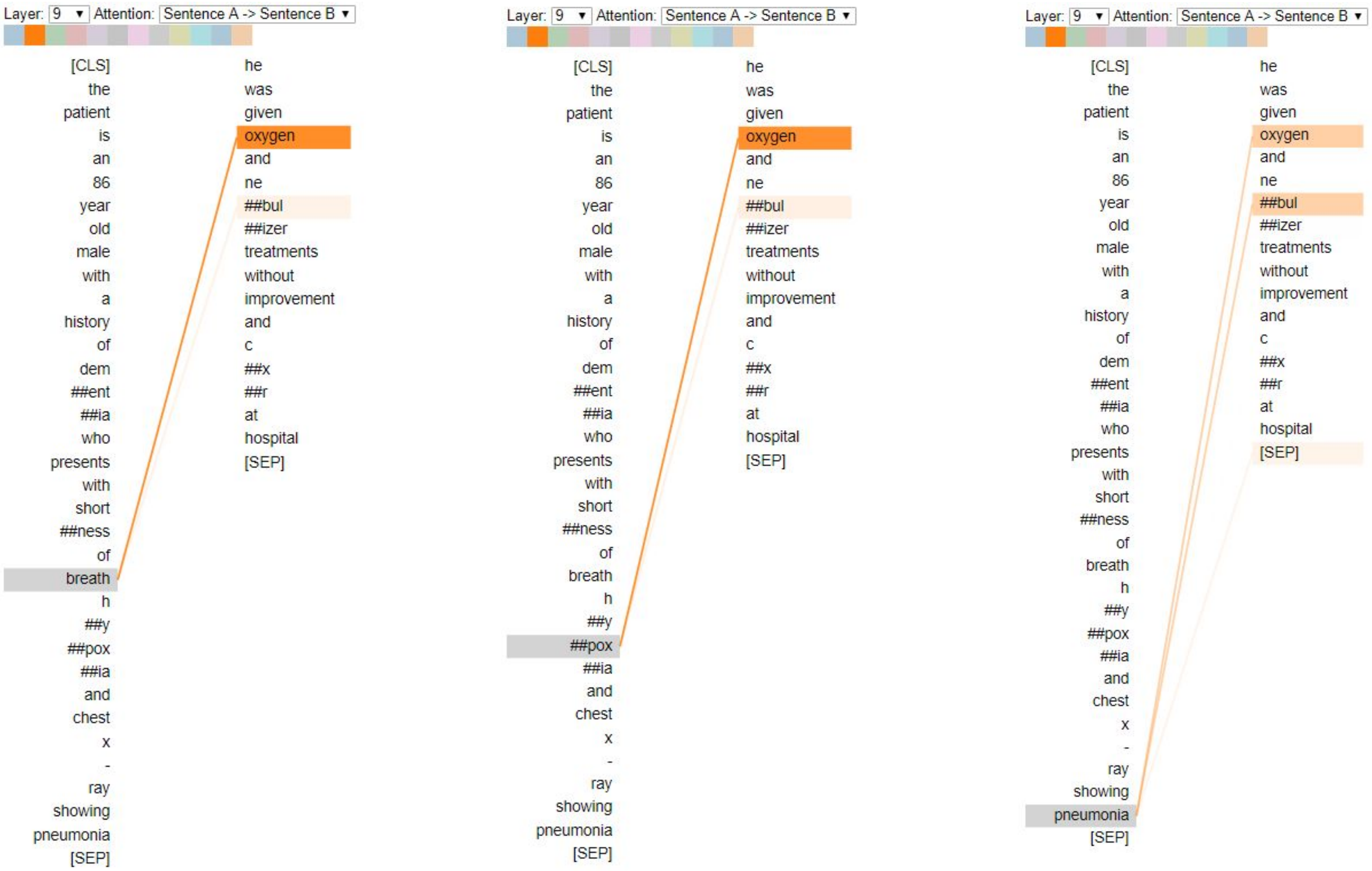
Understanding Symptoms/ Disease and Relation with Treatment

- Sentence A : The patient is an 86 year old male with a history of dementia who presents with **shortness of breath**, **hypoxia**, and chest x-ray showing **pneumonia**
- Sentence B: He was given **oxygen** and **nebulizer** treatments without improvement and CXR at Hospital

***hypoxia is the deficiency of oxygen in tissues

***nebulizer turns medicine from liquid to gas to directly reach the lungs


Layer: 0 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]
Layer: 1 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]
Layer: 9 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]
Layer: 10 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]

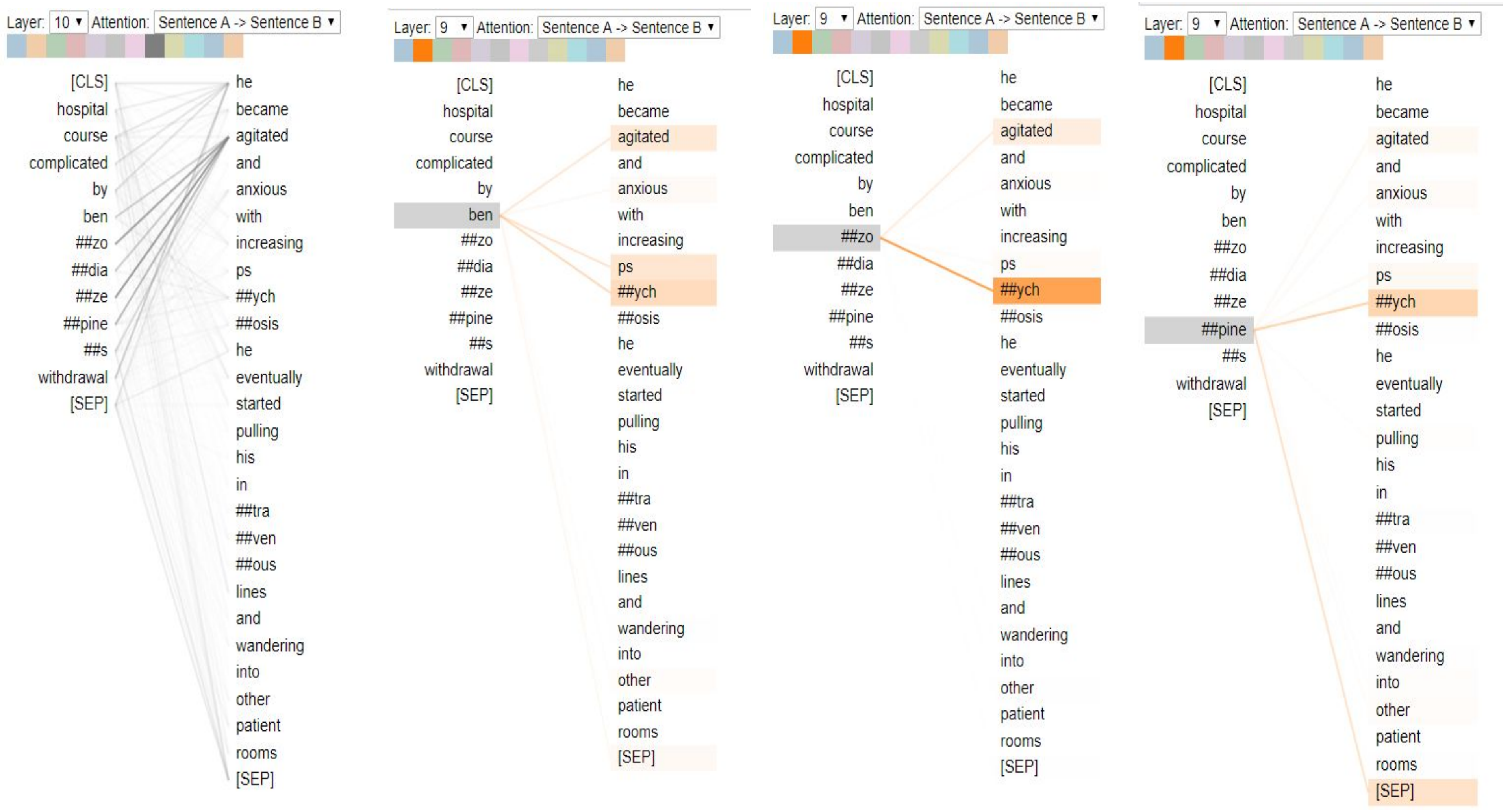


Layer: 9 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]
Layer: 9 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]
Layer: 9 Attention: Sentence A -> Sentence B
[CLS] the patient is an 86 year old male with a history of dem ##ent ##ia who presents with short ##ness of breath h ##y ##pox ##ia and chest x - ray showing pneumonia [SEP]
he was given oxygen and ne ##bul ##izer treatments without improvement and c ##x ##r at hospital [SEP]

Understanding drug's relation to symptoms, treatment, and side effects

- Sentence A : Hospital course complicated by **benzodiazepines withdrawal**
- Sentence B: He became **agitated** and **anxious** with increasing **psychosis**. He eventually started pulling his intravenous lines and wandering into other patient rooms

***benzodiazepines are sedatives that can cause psychosis in severe withdrawals and can be used calm aggravated patients



Layer: 10
Attention: Sentence A -> Sentence B
[CLS]
hospital
course
complicated
by
ben
##zo
##dia
##ze
##pine
##s
withdrawal
[SEP]
he
became
agitated
and
anxious
with
increasing
ps
##ych
##osis
he
eventually
started
pulling
his
in
##tra
##ven
##ous
lines
and
wandering
into
other
patient
rooms
[SEP]
Layer: 9
Attention: Sentence A -> Sentence B
[CLS]
hospital
course
complicated
by
ben
##zo
##dia
##ze
##pine
##s
withdrawal
[SEP]
he
became
agitated
and
anxious
with
increasing
ps
##ych
##osis
he
eventually
started
pulling
his
in
##tra
##ven
##ous
lines
and
wandering
into
other
patient
rooms
[SEP]
Layer: 9
Attention: Sentence A -> Sentence B
[CLS]
hospital
course
complicated
by
ben
##zo
##dia
##ze
##pine
##s
withdrawal
[SEP]
he
became
agitated
and
anxious
with
increasing
ps
##ych
##osis
he
eventually
started
pulling
his
in
##tra
##ven
##ous
lines
and
wandering
into
other
patient
rooms
[SEP]
Layer: 9
Attention: Sentence A -> Sentence B
[CLS]
hospital
course
complicated
by
ben
##zo
##dia
##ze
##pine
##s
withdrawal
[SEP]
he
became
agitated
and
anxious
with
increasing
ps
##ych
##osis
he
eventually
started
pulling
his
in
##tra
##ven
##ous
lines
and
wandering
into
other
patient
rooms
[SEP]

Future Directions and Limitations

- Pruning and compression
  - Some of these attention layers are “useless”
  - Methods such as Deep Compression (using pruning, quantization, and huffman coding) greatly reduces model size
  - This helps with for deployment of large models

- XLNet has shown promise in power over BERT for certain tasks

- Text summarization algorithms may help in focusing model’s attention on important information