Explainable Prediction of Medical Codes from Clinical Text

James Mullenbach
Sarah Wiegreffe
Jon Duke
Jimeng Sun
Jacob Eisenstein

The Clinical Coding Problem

- Electronic Health Records have seen increasing adoption in the last 5 years

The Clinical Coding Problem

- Electronic Health Records have seen increasing adoption in the last 5 years
- They abound with written physician notes

The Clinical Coding Problem

- Electronic Health Records have seen increasing adoption in the last 5 years
- They abound with written physician notes
- ICD: taxonomy of diagnoses and procedures

The Clinical Coding Problem

- Electronic Health Records have seen increasing adoption in the last 5 years
- They abound with written physician notes
- ICD: taxonomy of diagnoses and procedures
- Human coding laborious, error-prone [Birman-Deych et al., 2005]

The Clinical Coding Problem, as an NLPer

- Highly multi-*label* classification
  - Of nearly 9,000 labels, predict a subset

The Clinical Coding Problem, as an NLPer

- Highly multi-*label* classification
  - Of nearly 9,000 labels, predict a subset
- Testbed for document representations

The Clinical Coding Problem, as an NLPer

- Highly multi-*label* classification
  - Of nearly 9,000 labels, predict a subset
- Testbed for document representations
- Documents are long and loosely structured

The MIMIC dataset

- Open-access, de-identified

The MIMIC dataset

- Open-access, de-identified
- 47k admissions -> 47k documents for training

The MIMIC dataset

- Open-access, de-identified
- 47k admissions -> 47k documents for training
- **Loosely structured:**

```
Admission Date:  [**2118-6-2**]       Discharge Date:  [**2118-6-14**]

Date of Birth:                    Sex:  F

Service:  MICU and then to [**Doctor Last Name **] Medicine

HISTORY OF PRESENT ILLNESS:  This is an 81-year-old female
with a history of emphysema (not on home O2), who presents...
```

The MIMIC dataset

- Open-access, de-identified
- 47k admissions -> 47k documents for training
- **Loosely structured:**

```
Admission Date:  [**2118-6-2**]       Discharge Date:  [**2118-6-14**]

Date of Birth:                    Sex:  F

Service:  MICU and then to [**Doctor Last Name **] Medicine

HISTORY OF PRESENT ILLNESS:  This is an 81-year-old female
with a history of emphysema (not on home O2), who presents...
```

Long: Median post-processed document length: 1,341

The MIMIC dataset

- Open-access, de-identified
- 47k admissions -> 47k documents for training
- **Loosely structured:**

```
Admission Date:  [**2118-6-2**]       Discharge Date:  [**2118-6-14**]

Date of Birth:                    Sex:  F

Service:  MICU and then to [**Doctor Last Name **] Medicine

HISTORY OF PRESENT ILLNESS:  This is an 81-year-old female
with a history of emphysema (not on home O2), who presents...
```

Many labels:

```
519.1: 'Other disease…'
491.21: 'Obstructive …'
518.81: 'Acute respir…'
486: 'Pneumonia, orga…'
276.1: 'Hyposmolality…'
244.9: 'Unspecified h…'
31.99: 'Other operati…'
.
.
.
```

Long: Median post-processed document length: 1,341

Median # labels: 14

Prior approaches

- Predict from a subset of labels

This work

- Predict from all labels

Prior approaches

- Predict from a subset of labels
- Sub-domain focus, e.g. radiology

This work

- Predict from all labels
- General ICU setting

Prior approaches

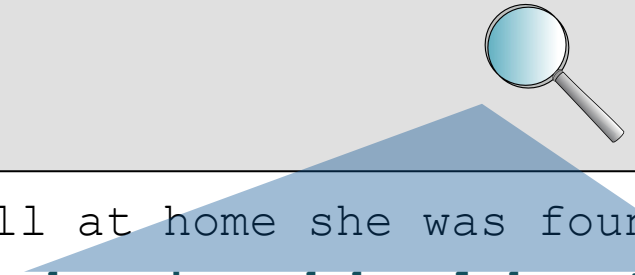
- Predict from a subset of labels
- Sub-domain focus, e.g. radiology
- Private dataset

This work

- Predict from all labels
- General ICU setting
- Open-access data

Modeling desiderata

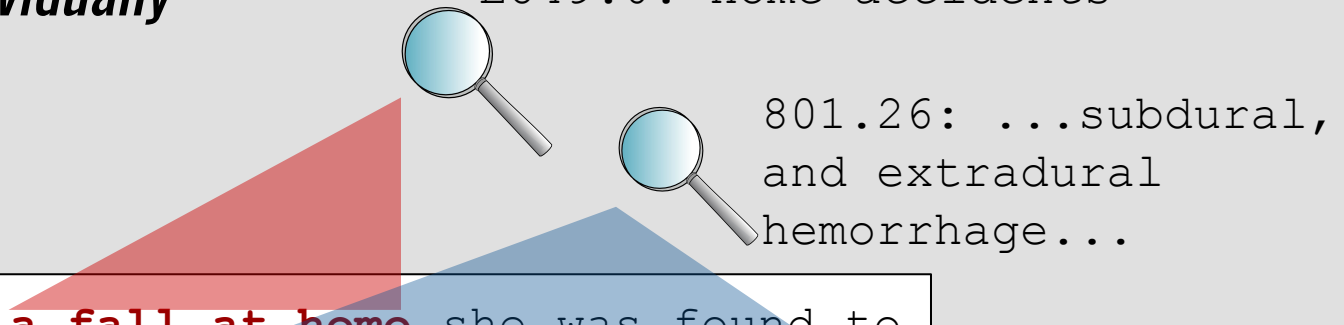
- Focus on the parts that matter



...who sustained a fall at home she was found to have a large acute on **chronic subdural hematoma** with extensive midline shift...

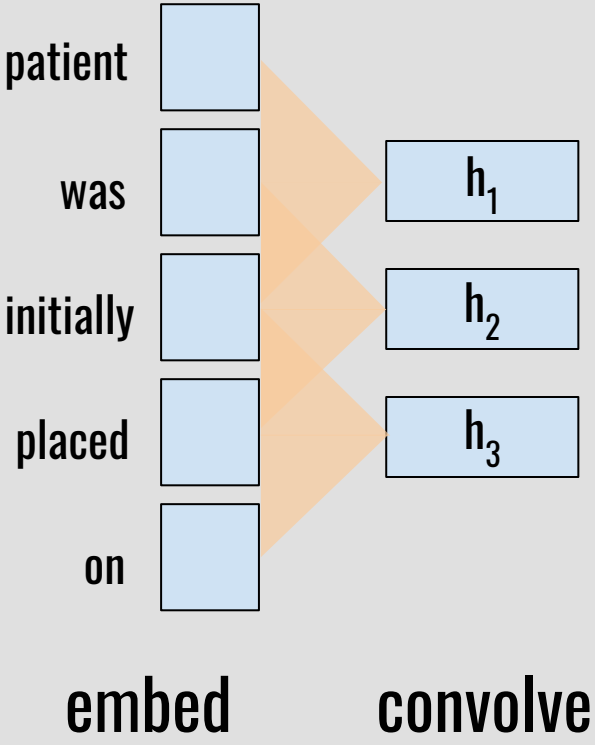
Modeling desiderata

- Focus on the parts that matter
- Treat labels *individually*

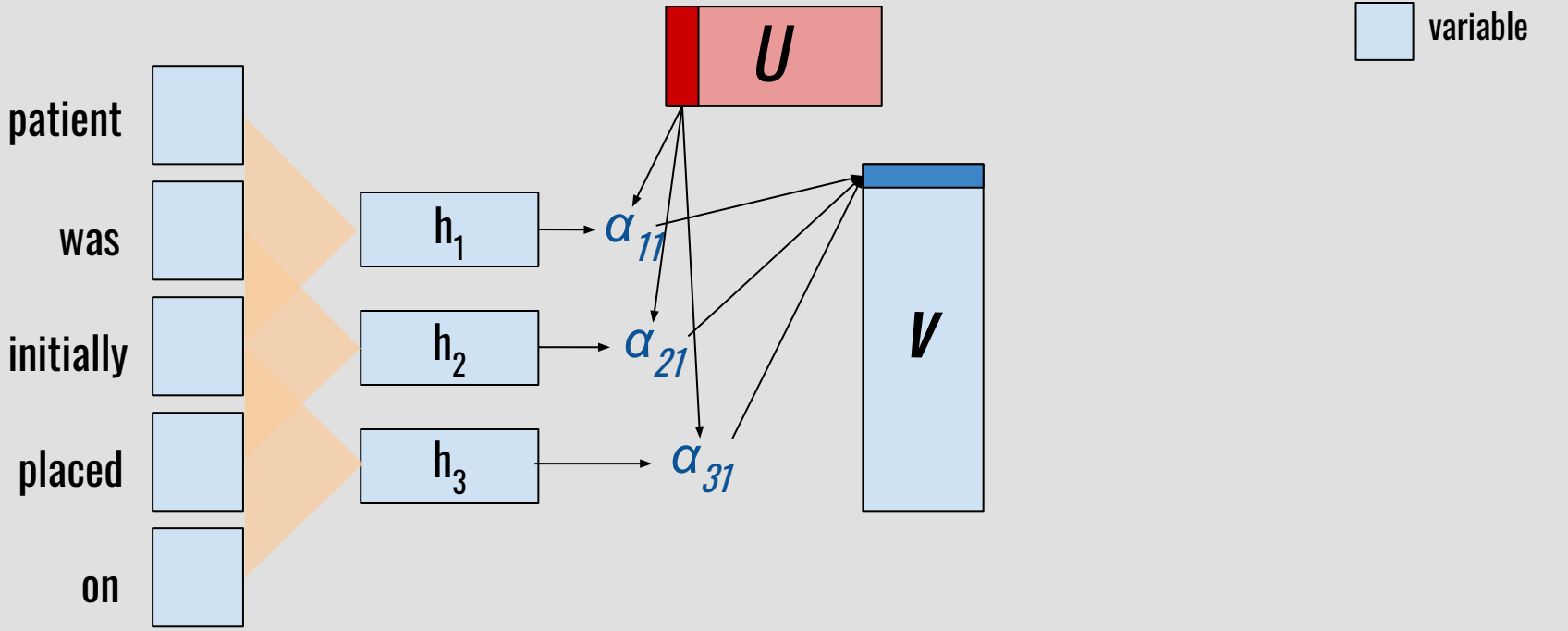


...who sustained **a fall at home** she was found to have a large acute on **chronic subdural hematoma** with extensive midline shift...

Convolution

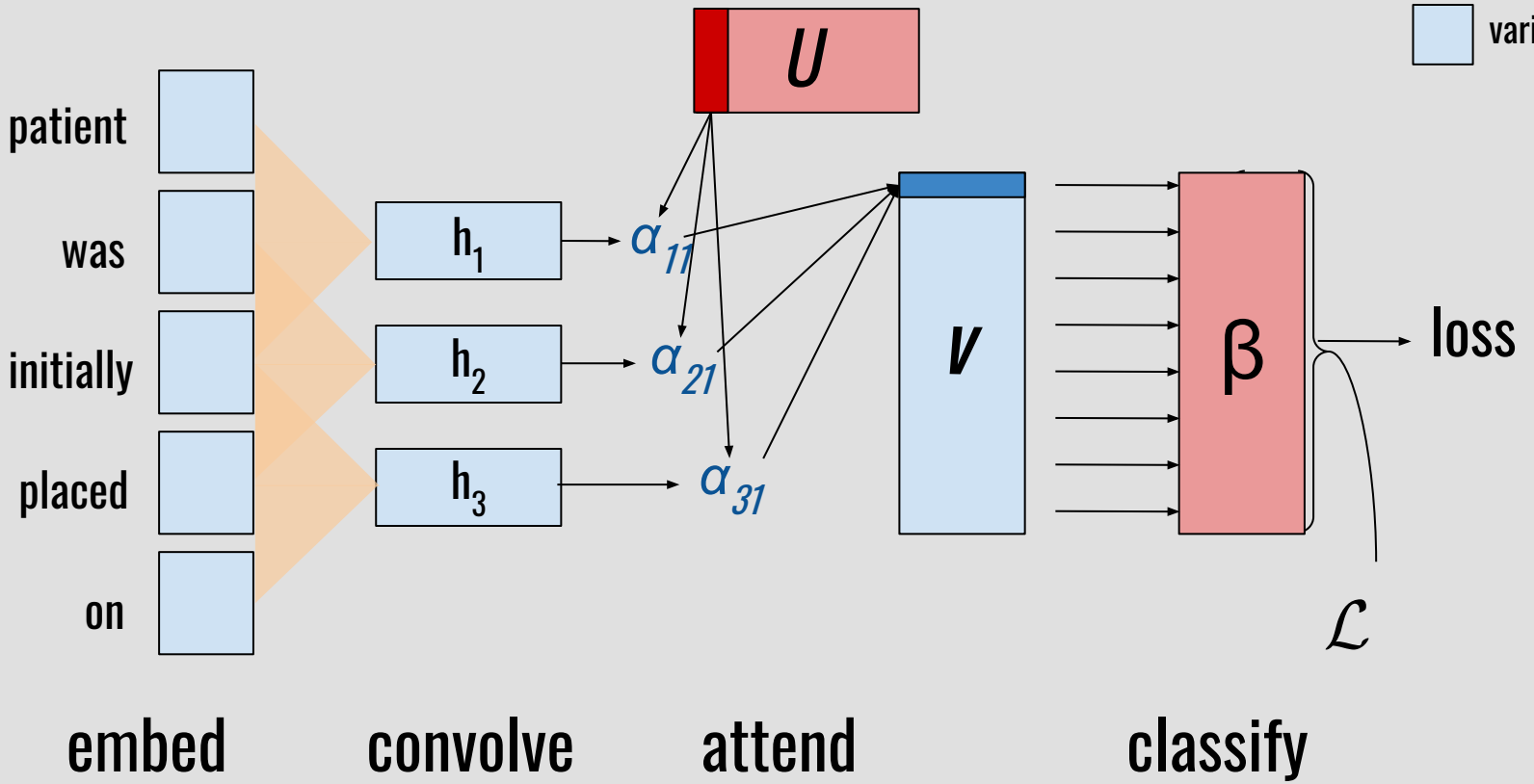


Convolution + Attention



embed convolve attend (separately for each label)

Convolution + Attention = the CAML model



Dealing with the long tail

- Huge label space (nearly 9,000 total)

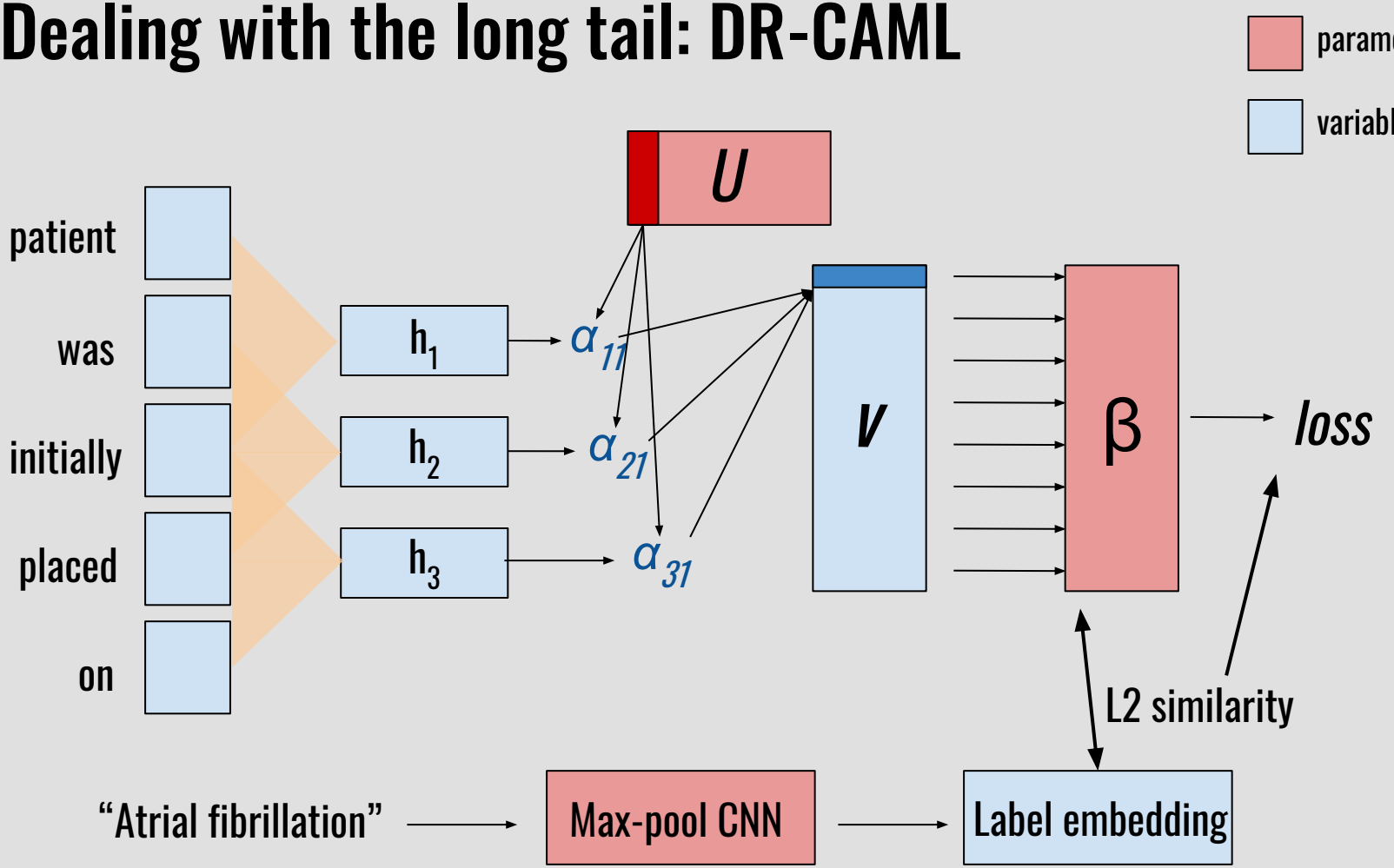
Dealing with the long tail

- Huge label space (nearly 9,000 total)
- Many labels are similar

250.00: “Diabetes mellitus without mention of complication, type II or unspecified type, not stated as uncontrolled”

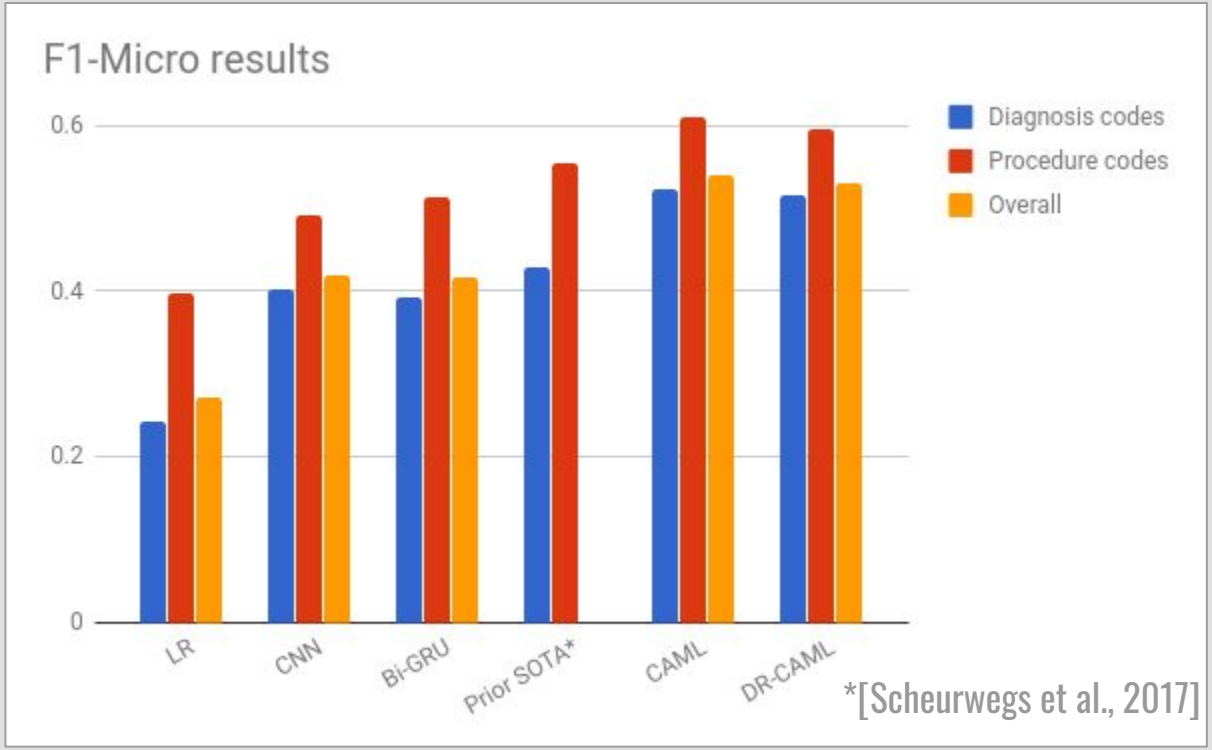
250.02: “Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled”

Dealing with the long tail: DR-CAML



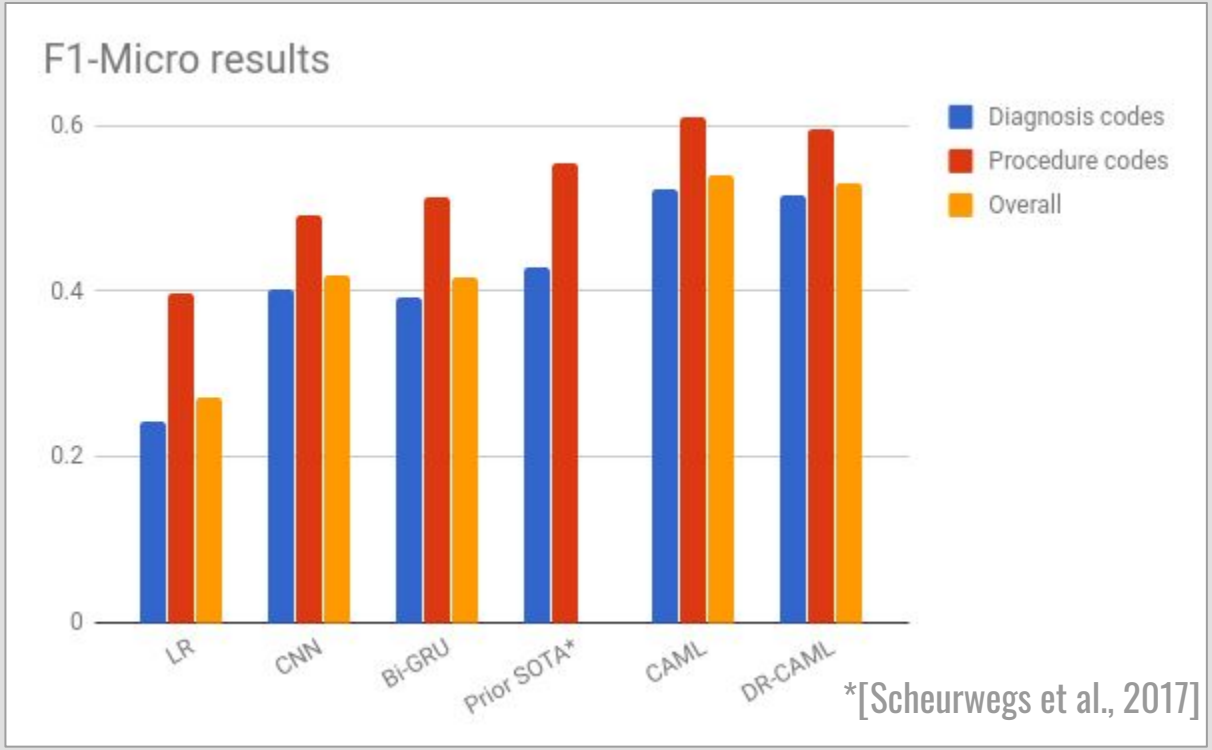
MIMIC results

- Few works use the whole label space



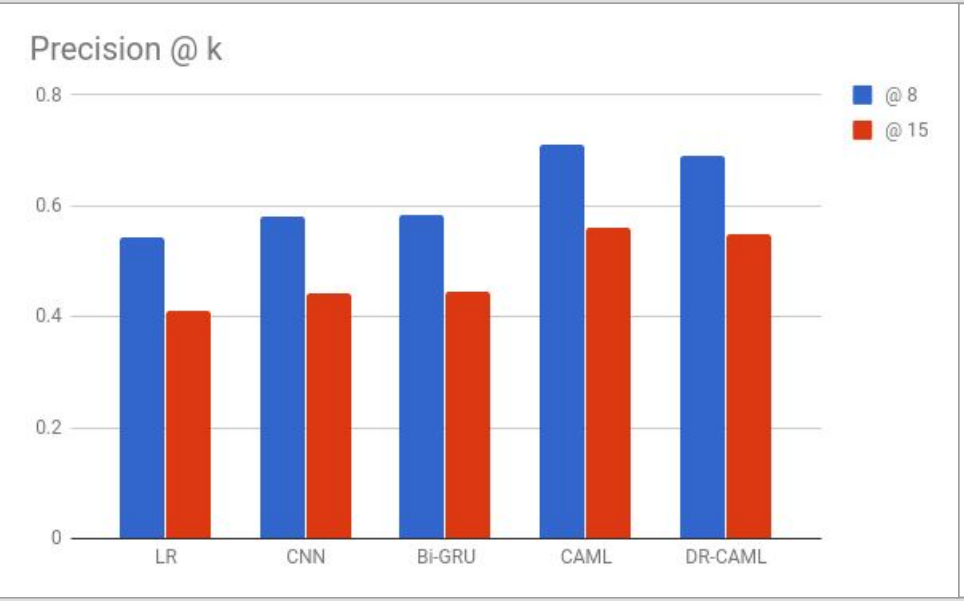
MIMIC results

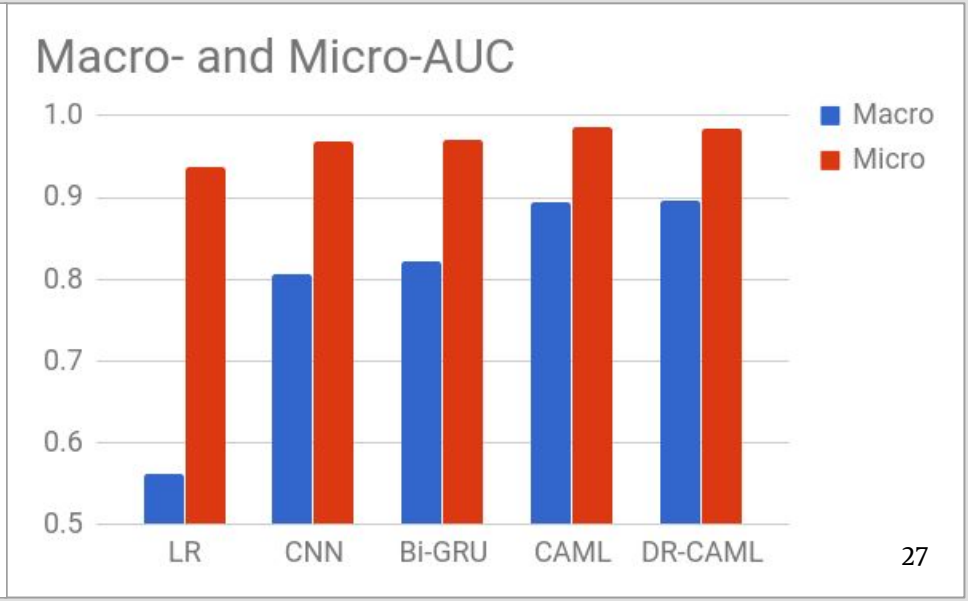
- Few works use the whole label space
- 20%, 8% relative improvement over prior SotA



More MIMIC metrics

- Enable future comparison
- Precision @ k: decision support use-case





But, is the model focusing on reasonable areas?

- We want interpretability, especially in the clinical domain

But, is the model focusing on reasonable areas?

- We want interpretability, especially in the clinical domain
- Physician evaluation: select all text snippet(s) that explain the code

But, is the model focusing on reasonable areas?

- We want interpretability, especially in the clinical domain
- Physician evaluation: select all text snippet(s) that explain the code
- Baselines: LogReg, CNN, cosine similarity

But, is the model focusing on reasonable areas?

- We want interpretability, especially in the clinical domain
- Physician evaluation: select all text snippet(s) that explain the code
- Baselines: LogReg, CNN, cosine similarity
- 100 random label-document samples

But, is the model focusing on reasonable areas?

- We want interpretability, especially in the clinical domain
- Physician evaluation: select all text snippet(s) that explain the code
- Baselines: LogReg, CNN, cosine similarity
- 100 random label-document samples
- “Informative” and “Highly Informative”

Physician evaluation example*

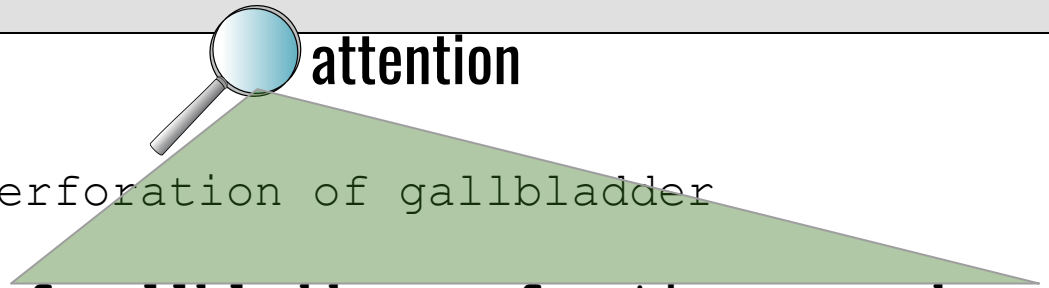
Code: 575.4

Full descriptions: Perforation of gallbladder

“. . in the setting of gallbladder perforation secondary to acute acalculous cholecystitis after . . . . . inhalation hospital1 times a day metronidazole mg tablet sig one tablet po tid times . . . . . to have an infection in your gallbladder requiring iv antibiotics and tube placement for . . .”

*not exact format used

Physician evaluation example*



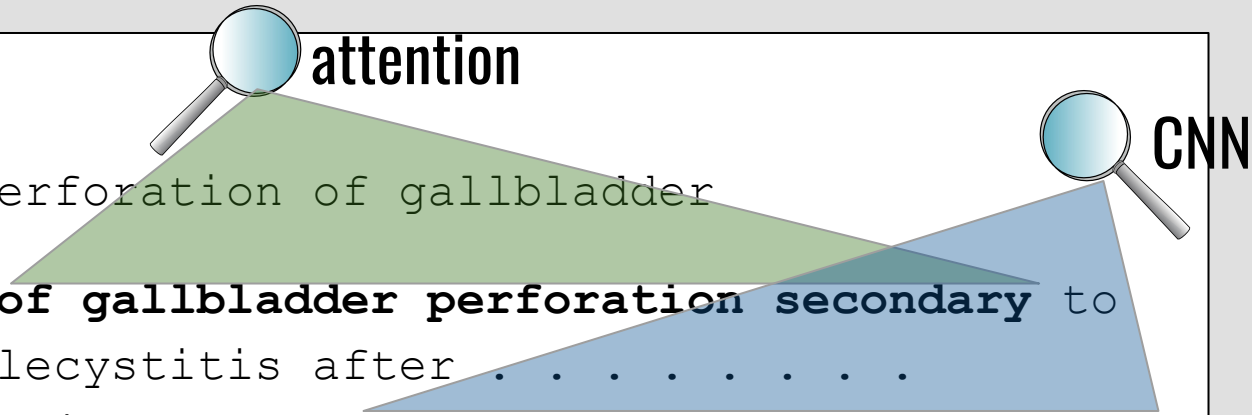
Code: 575.4

Full descriptions: Perforation of gallbladder

“. . in the setting **of gallbladder perforation secondary** to acute acalculous cholecystitis after . . . . . inhalation hospital1 times a day metronidazole mg tablet sig one tablet po tid times . . . . . to have an infection in your gallbladder requiring iv antibiotics and tube placement for . . .”

*not exact format used

Physician evaluation example*



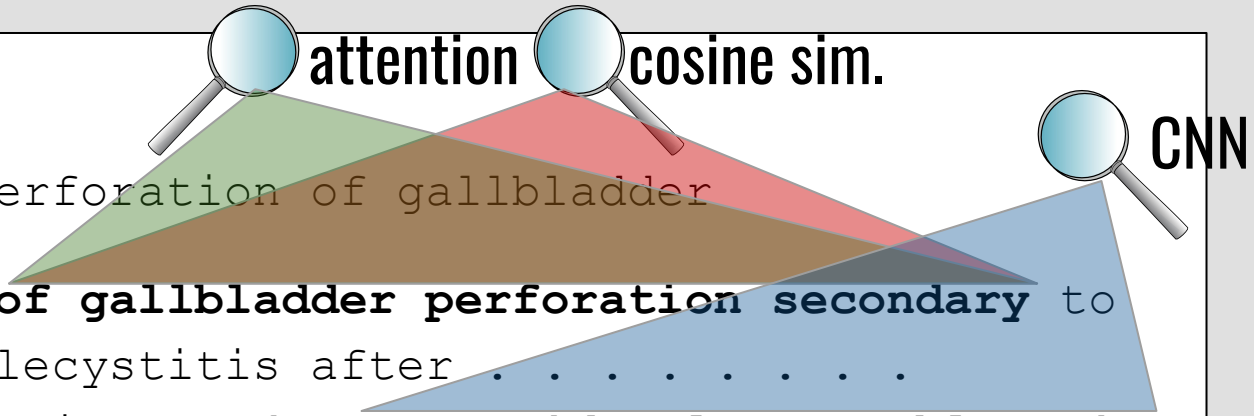
Code: 575.4

Full descriptions: Perforation of gallbladder

". . in the setting **of gallbladder perforation secondary** to acute acalculous cholecystitis after . . . . . inhalation hospital1 times a day **metronidazole mg tablet sig** one tablet po tid times . . . . . to have an infection in your gallbladder requiring iv antibiotics and tube placement for . . ."

*not exact format used

Physician evaluation example*



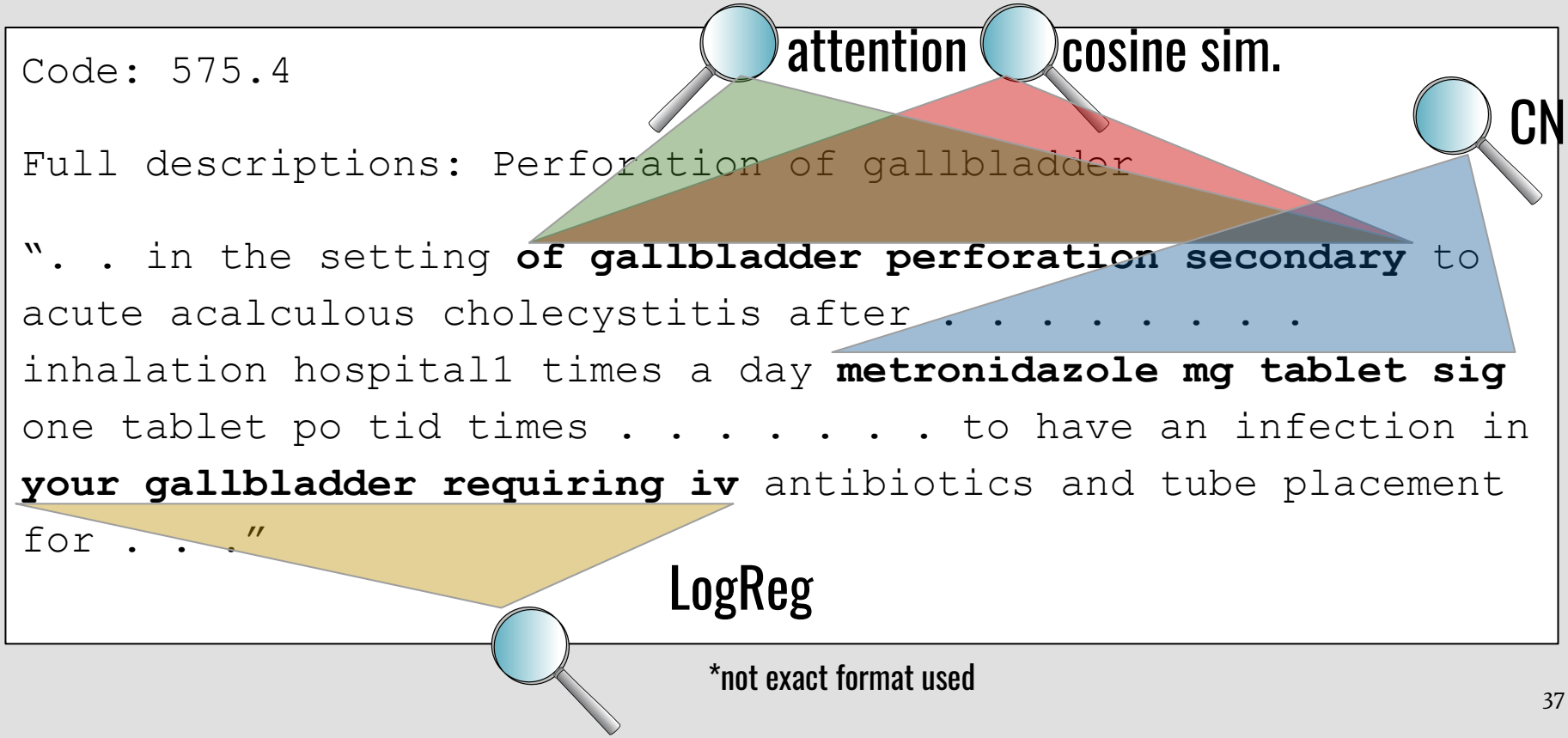
Code: 575.4

Full descriptions: Perforation of gallbladder

“. . in the setting **of gallbladder perforation secondary** to acute acalculous cholecystitis after . . . . . inhalation hospital1 times a day **metronidazole mg tablet sig** one tablet po tid times . . . . . to have an infection in your gallbladder requiring iv antibiotics and tube placement for . . .”

*not exact format used

Physician evaluation example*



Code: 575.4

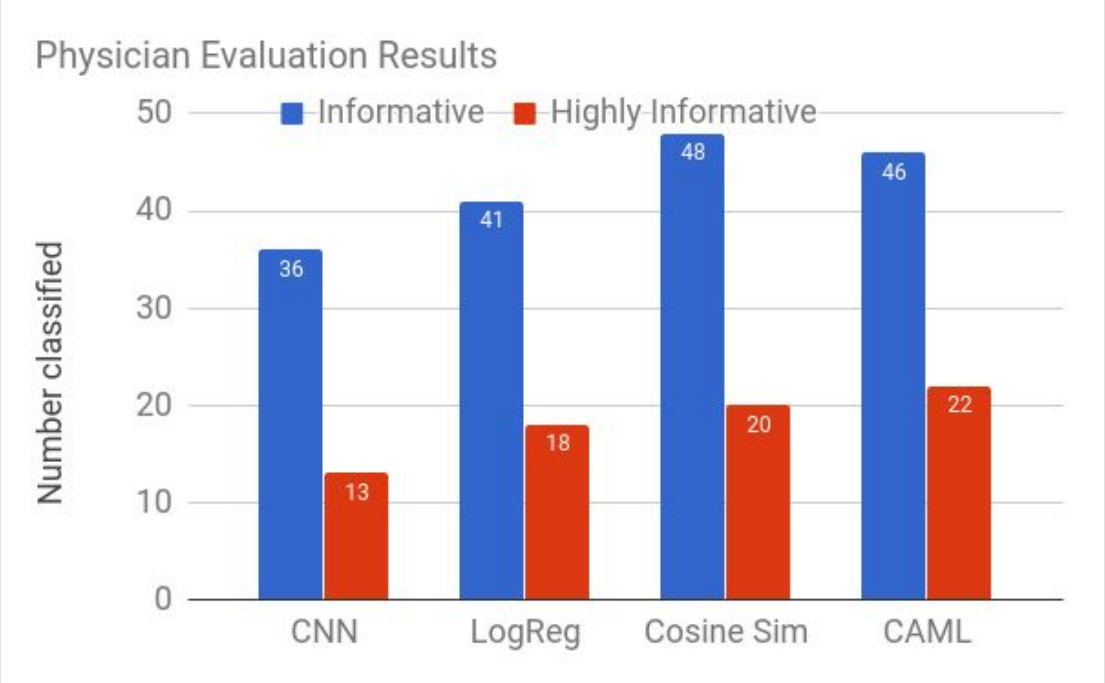
Full descriptions: Perforation of gallbladder

“. . in the setting **of gallbladder perforation secondary** to acute acalculous cholecystitis after . . . . . inhalation hospital1 times a day **metronidazole mg tablet sig** one tablet po tid times . . . . . to have an infection in **your gallbladder requiring iv** antibiotics and tube placement for . . .”

*not exact format used

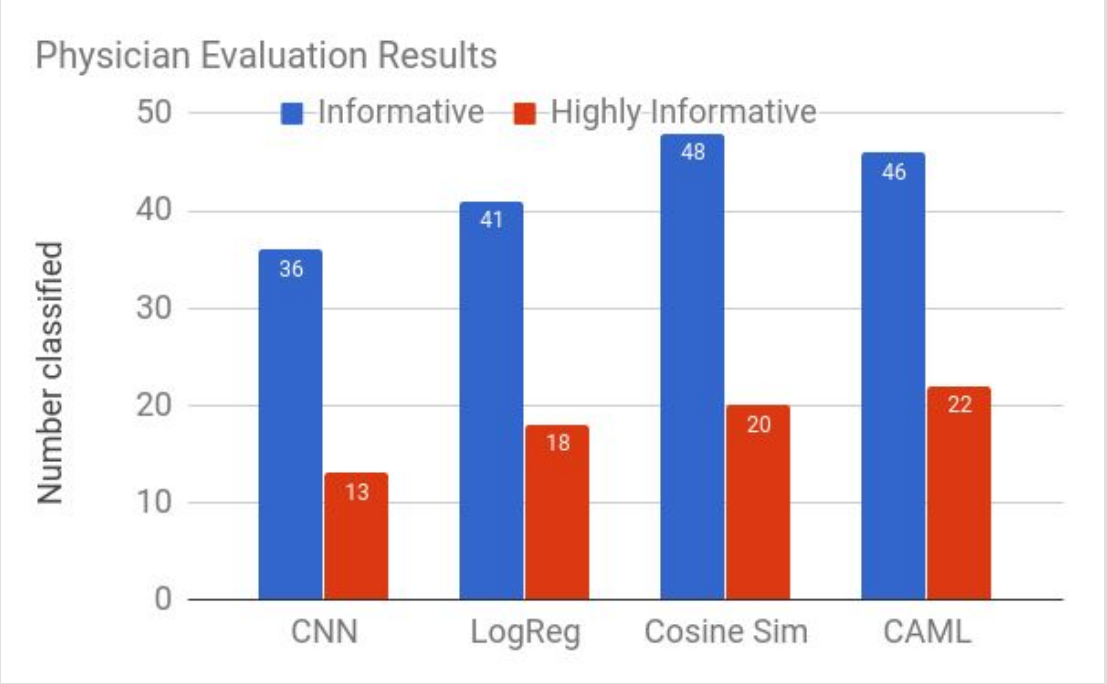
Physician evaluation results

- Improves upon CNN, LogReg



Physician evaluation results

- Improves upon CNN, LogReg
- More experts needed!



Future work

- Exploit the loose structure of discharge summaries
  - Some have already done this [Shi et al., 2017]

Future work

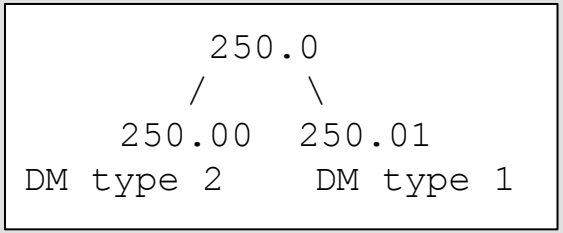
- Exploit the loose structure of discharge summaries
  - Some have already done this [Shi et al., 2017]
- Sparsify or structure attention to improve interpretability, accuracy
  - “fusedmax” [Niculae & Blondel, 2017]

Future work

- Exploit the loose structure of discharge summaries
  - Some have already done this [Shi et al., 2017]
- Sparsify or structure attention to improve interpretability, accuracy
  - “fusedmax” [Niculae & Blondel, 2017]
- Develop methods for few-shot or zero-shot classification

Future work

- Exploit the loose structure of discharge summaries
  - Some have already done this [Shi et al., 2017]
- Sparsify or structure attention to improve interpretability, accuracy
  - “fusedmax” [Niculae & Blondel, 2017]
- Develop methods for few-shot or zero-shot classification
- Exploit hierarchy of codes
  - GRAM [Choi et al., 2017]



Summary

- ICD coding is valuable and challenging

Summary

- ICD coding is valuable and challenging
- Convolution + attention works well

Summary

- ICD coding is valuable and challenging
- Convolution + attention works well
- Attention can explain the predictions

Summary

- ICD coding is valuable and challenging
- Convolution + attention works well
- Attention can explain the predictions

Thank you!

Questions?

Code, data splits, pretrained models: github.com/jamesmullenbach/caml-mimic