Prediction of Psychiatric Conditions Based on "History of Present Illness"

Tung Tran and Ramakanth Kavuluru

Division of Biomedical Informatics, Internal Medicine
Department of Computer Science
University of Kentucky

CEGS N-GRID 2016 i2b2 NLP Novel Data Use Track

Outline

1. Motivation
2. Methods
3. Results

Psychiatric Notes are Rich in Information

- Many narratives and labeled fields, some appearing consistently
- Each note instance varies greatly in length and scope
- Data exists in the form of non-machine readable, free-form text
- Formatting issues and inconsistencies make parsing tricky
- Spelling errors can introduce noise in predictive model

Problem Statement

- We have short narratives about a patient's History of Present Illness and Precipitating Events – average length 300 words.
- We have labels indicating presence of psychological disorders in the Psychiatric Review of Systems portion.

Can we predict a mental disorder using history of present illness?

Problem Statement

- We have short narratives about a patient's History of Present Illness and Precipitating Events – average length 300 words.
- We have labels indicating presence of psychological disorders in the Psychiatric Review of Systems portion.

Can we predict a mental disorder using history of present illness?

Problem Statement

- We have short narratives about a patient's History of Present Illness and Precipitating Events – average length 300 words.

- We have labels indicating presence of psychological disorders in the Psychiatric Review of Systems portion.

Can we predict a mental disorder using history of present illness?

Real-world Applications

- Allows for quick assessments based on a short narrative – can lead to early hints of a mental disorder
- Can be used to assist psychiatrists and health professionals in filling out the corresponding fields in forms
- Allows for automated surveillance of a patient's ongoing mental condition, possibly on a large scale

Making Predictions Based on a Short Narrative

- An example of "History of Present Illness"

  41 y/o single right-handed man, currently employed in construction, history of polysubstance use (etoh, cocaine, MJ), asthma, seasonal allergies and hypothyroidism, prior left shoulder and right ankle and knee surgeries s/p ortho trauma, treated with abilify for mood stabilization (no clear hx of bipolar disorder, psychotic disorder), here for psychopharmacology appt.

- Direct signals: No clear hx of bipolar disorder, psychotic disorder

- Indirect signal: Abilify (Antipsychotic)

Making Predictions Based on a Short Narrative

- Another example of “History of Present Illness”

  Prior episode of depression and treated with Effexor 8 years ago when she got married. Divorced 3 years later and she felt better, so went off. Marriage was abusive.

- Direct signal: Depression
- Indirect signal: Effexor (antidepressant)
- Note the variance in content scope and length

Preprocessing

1. Fix formatting issues with the data
   - Remove line-breaks in inappropriate places
   - Add line-breaks where missing
   - Fix frequent spelling errors and causes of noise
2. Parse each field into a key-value dictionary using regular expression patterns

Target of Label Prediction

- Many responses refer to a specific case of a disorder. E.g.

  DEMENTIA: Has anyone told the patient they are concerned the patient has memory problems: No

  DEMENTIA: Does the patient have trouble learning new information: No

- We model at a coarser level
  - Responses for a disorder are collapsed into a single label to keep it simple and to minimize class imbalance
  - E.g. The label for DEMENTIA is Yes if either sub-response is Yes.

Target of Label Prediction

Final list of 11 psychological disorders chosen for prediction

- ADHD
- Anxiety
- Bipolar
- Dementia
- Depression
- Eating Disorder
- Grief
- OCD/OCSD
- Panic
- Psychosis
- PTSD

Other labels of interest

- We also try to predict on other labels that are equally interesting:
  - Whether the patient has a history of
    1. Suicidal Behavior
    2. Inpatient Treatment
    3. Drug use
  - Additionally,
    4. Current smoking status

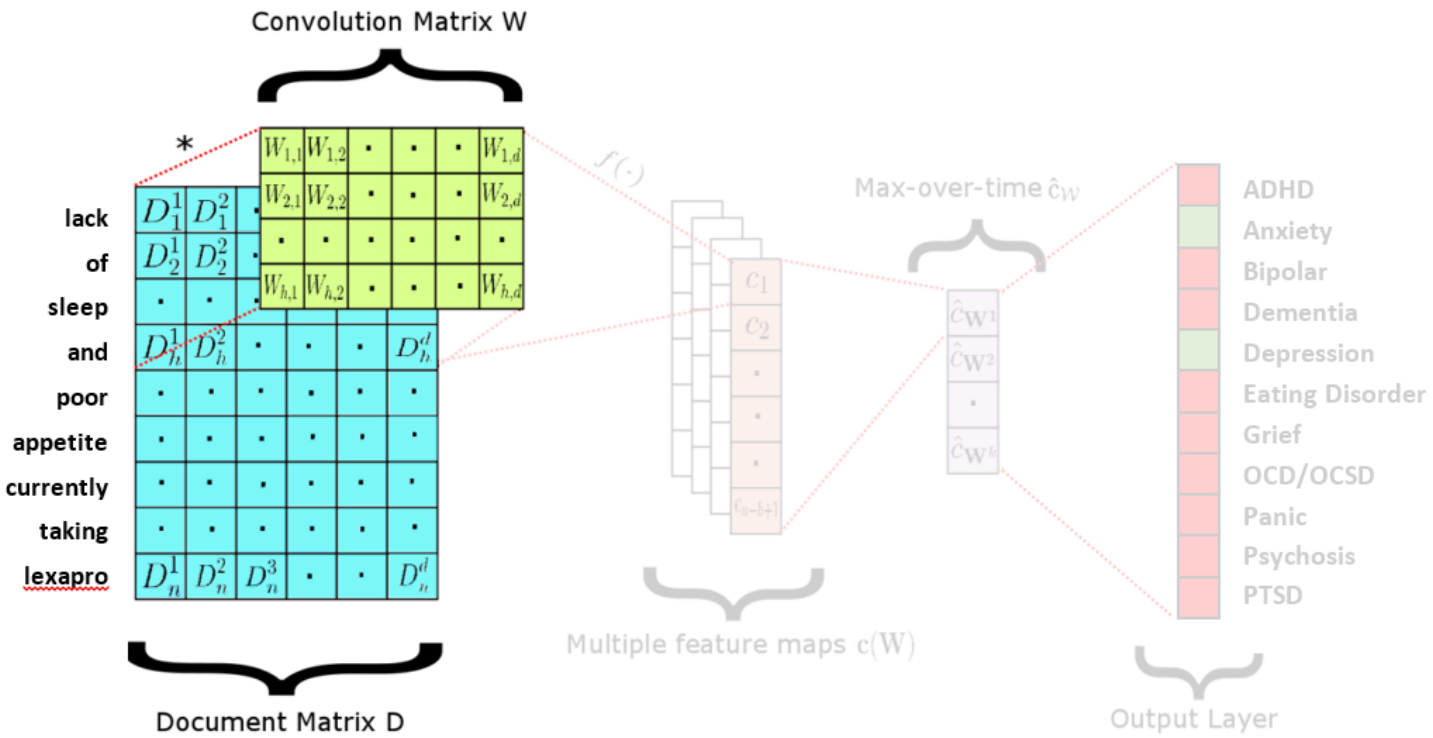
Methods for Prediction

- Support Vector Machines (SVM)
- Named Entity Recognition and Concept Mapping (NER)
- Convolutional Neural Network (CNN)
  - With Meta-Labeler
  - With Optimized-Thresholding

MetaMap - Name Entity Recognition (NER)

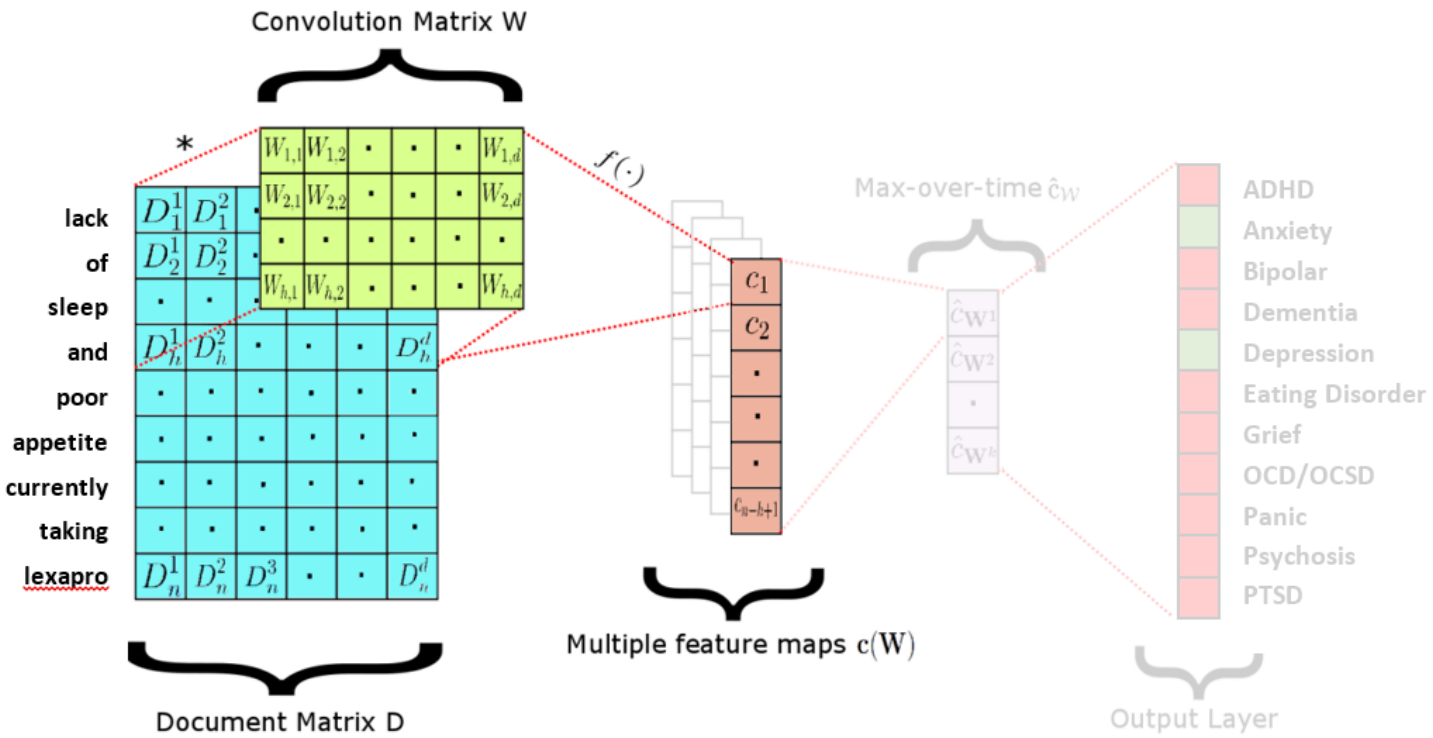
- A set of related CUIs is hand curated for each label using UMLS Metathesaurus as a reference.
- For each instance, we run NLM's MetaMap to extract its concepts.
- We predict Yes for an instance I and some target label L, if and only if $CuratedCUIs(L) \cap MetaMapCUIs(I) \neq \emptyset$.
- This method serves as an important baseline for comparison.

Convolutional Neural Network (CNN) Model Layout



- Input is a document matrix where each row corresponds to a word in the sequence and is represented by its word vector

Convolutional Neural Network (CNN) Model Layout



- Apply 2-D convolution over the document matrix to get a corresponding feature map

Convolutional Neural Network (CNN) Model Layout

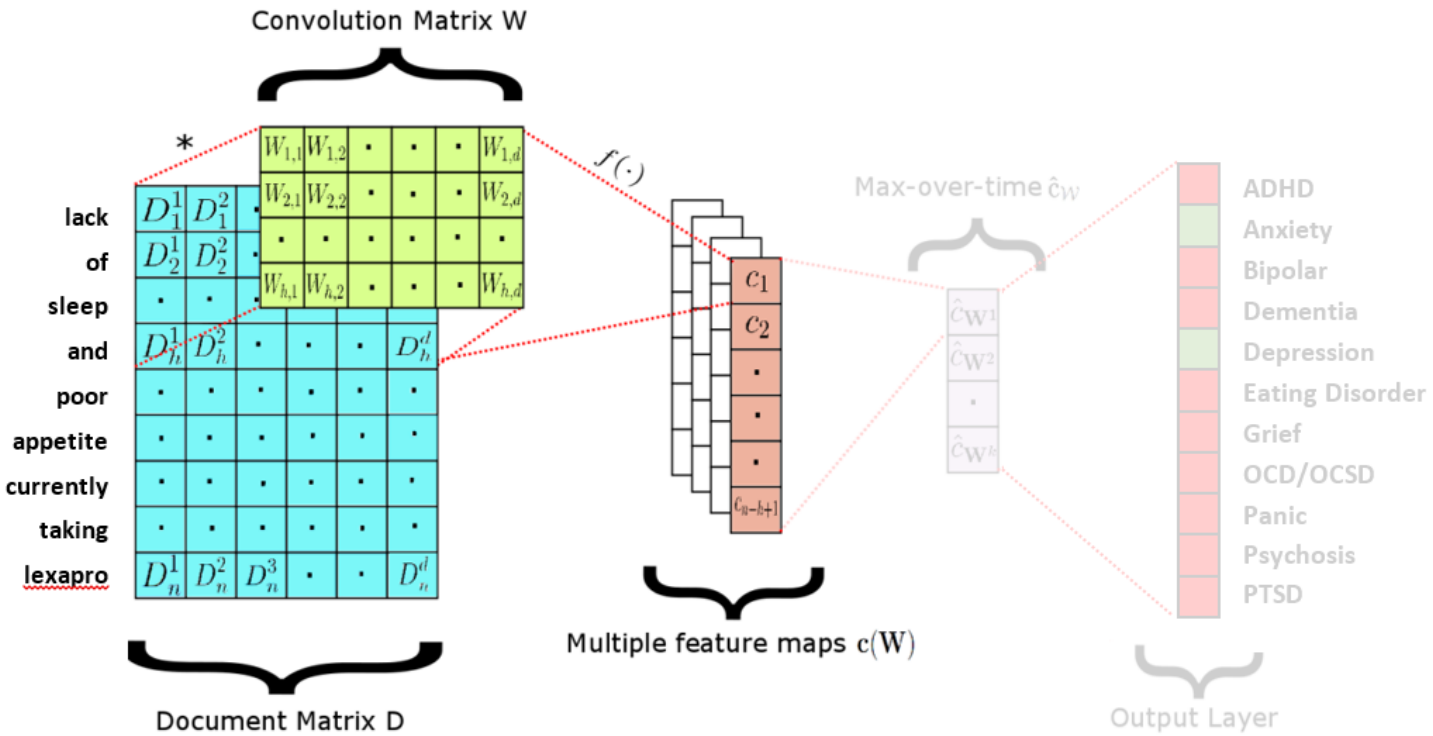
- A convolution operation takes as input a document matrix D and a convolution filter W of size h

$$W * D_{j:j+h-1} = \sum_{i=j}^{j+h-1} \sum_{k=0}^{d-1} W_{i,k} D_{i,k} \tag{1}$$

$$c_j = f(W * D_{j:j+h-1} + b) \tag{2}$$

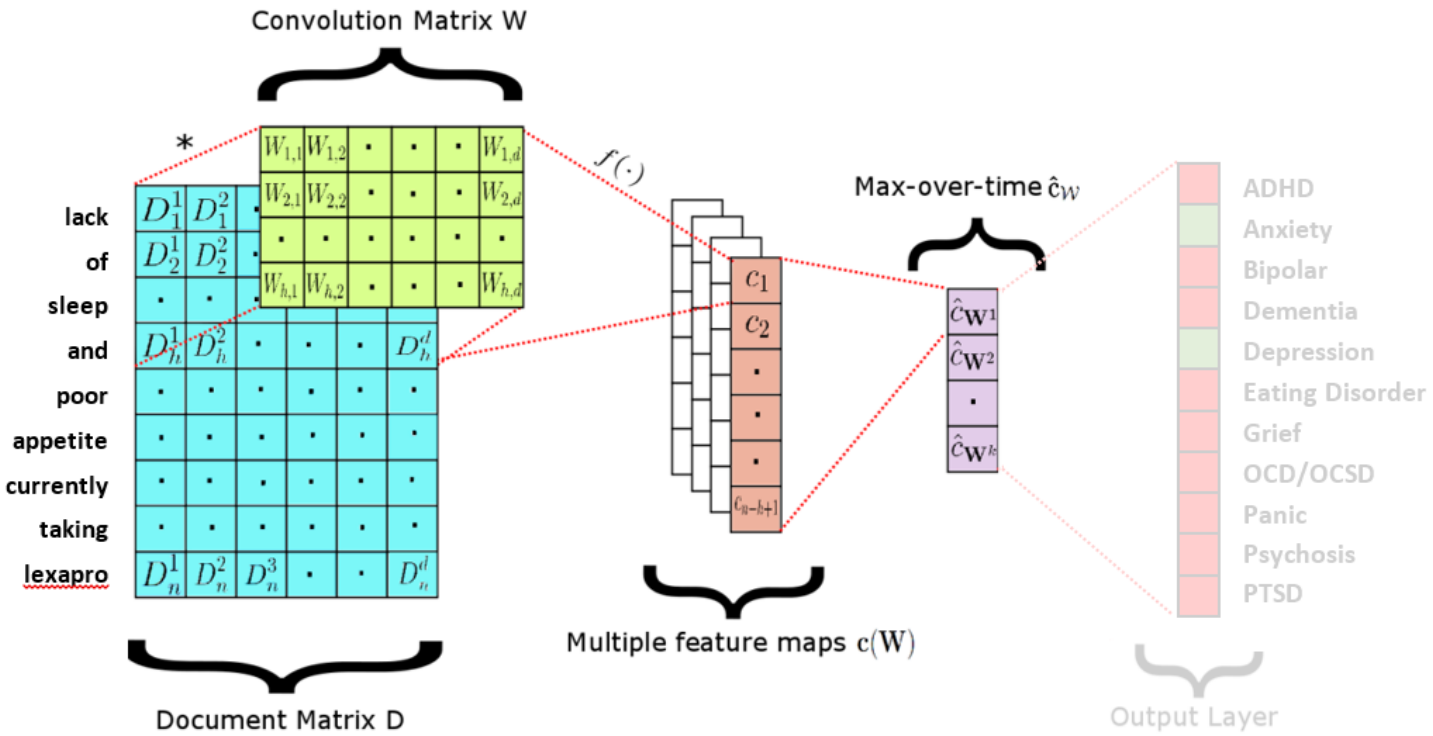
- Where f is a non-linear activation function and b is a bias term. We use a linear rectifier $f(x) = max(0, x)$.

Convolutional Neural Network (CNN) Model Layout



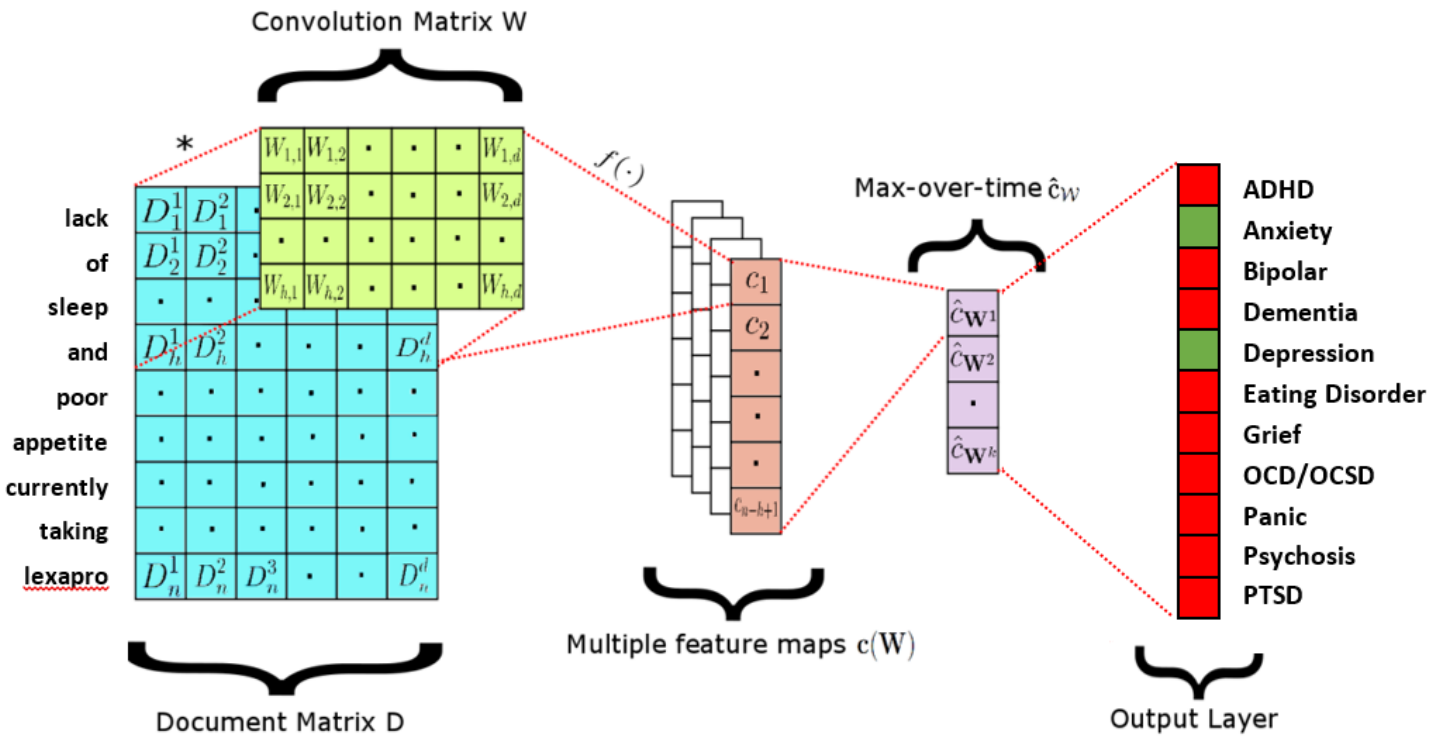
- The convolution operation produces a feature map.
- We can learn k convolution filters corresponding to k feature maps

Convolutional Neural Network (CNN) Model Layout



- Next we apply max-pooling over the k feature maps to produce the final feature vector

Convolutional Neural Network (CNN) Model Layout



- The output layer is fully-connected to the feature vector and contains one unit sigmoid unit per label.
- Each sigmoid unit produces an output $\in [0, 1]$ which can be used as a probability estimate.

CNN Variants

- CNN base
  - Predict Yes if the output is greater than 0.5 for some label.
- CNN with Meta-Labeler
  - Labels are ranked by their output scores and the top-k labels are predicted as Yes.
  - The value k is a hyper-parameter predicted for each instance by a separate linear regression model, which is trained on uni-gram/bi-gram features.
- CNN with Optimized-Thresholding
  - A customized threshold is chosen for each label instead of 0.5
  - Each per-label threshold is optimized separately based on the validation fold

Training

- We performed 10-fold cross-validation on all released notes
- We trained on first 8 folds, fine-tuned hyper-parameters on 9th fold, and tested on 10th fold
- With regard to CNNs,
  - We used convolution filters of sizes 3, 4, and 5
  - For each filter size, we learned 250 convolution filters
  - Final predictions are made by averaging the outputs scores of 10 CNNs trained with different random parameter initializations

Overall Evaluation

	SVM	NER	CNN	CNN/Meta-Labeler	CNN/Thresholding
Micro-Precision (%)	51.6	**67.9**	64.4	65.3	64.0
Micro-Recall (%)	**64.3**	33.2	58.1	50.7	54.1
Micro-F1 (%)	57.3	44.6	**61.1**	57.1	58.7

Table: Results comparing overall effectiveness

- Overall, the base CNN method achieves the best F1-score with a good balance of precision and recall.
- NER results in very high precision but low recall for depression and other conditions, likely because it is looking for very specific terms in the text with little inference.

Per-Label Evaluation

- History of present illness is a good predictor of some conditions, such as depression and anxiety, with F-scores of 87 (SVM) and 81 (CNN/Thresholding) respectively. However, these two are also the most frequent labels (occur in 77% and 68% of all records resp.)
- NER does very well in predicting depression with F-score of 83 since depression is more likely to be stated explicitly in the text.
- After depression and anxiety, we notice F-scores over 60 for ADHD, bipolar disorders, and panic attacks.

Per-Label Evaluation

- The best performing model overall (base CNN) is inferior to other models (except NER) in predicting presence of grief.
- For grief, the CNN/Meta-Labeler model provides a 25 point improvement over the base CNN model.
- Performance is poor overall for the non-psychiatric labels such as history of suicidal behavior and smoking status. Potentially not enough signal in the narrative to make a reliable prediction.

Summary

- We showed that a short “history of present illness” blurb can predict some psychiatric conditions.
- We built models using SVMs, NER, and CNNs and performed 10-fold cross-validation on all available notes.
- Outlook: Accommodate label correlations and fine-tune individual label classifiers, may lead to overall improvement in performance.

 Thank you 

This effort is supported by the National Library of Medicine, Kentucky Lung Cancer Research Program, and National Center for Advancing Translational Sciences