



@UUtah

## NLP-Assisted Differential Diagnosis of Chronic Obstructive Pulmonary Disease Exacerbation

---

Fatemeh Shah-Mohammadi, Assistant Professor

Joseph Finkelstein, Professor

*University of Utah*





## Introduction

---

- Benefits of early identification of COPD exacerbation:
  - Timely prescription of optimal treatment plan
  - Cost reduction by avoiding unnecessary tests and specialist consultations
- Aim of this study:
  - Leveraging NLP to extract features from clinical notes documented within 24 hours of admission to the hospital.
  - Conduct differential diagnosis in COPD patients by developing a classification model.



## Dataset

---

- Formed by a manual review of patient charts and confirming the diagnosis of COPD.
- Contained patients' socio-demographic information along with different notes documented during their admission.
- All patients in the cohort were admitted to the emergency department (ED) and then, depending on the severity of symptoms and clinical presentation, hospitalized for further treatment.
- Considered very first documented notes: ED triage notes and ED provider notes.



## Dataset

- Triage note described the reason for the patient's visit, including specific symptoms and incidents.
- ED provider note used for providing documentation of the patient assessment throughout the emergency department visit.
- Contained various information about the patients, including the very first vital signs observed in ED, such as blood pressure and respiratory rate.
- Used discharge summaries to label the patients (Manual review).

Labels	Number of patients (n=80)	%
1 - discharge diagnosis of COPD exacerbation	35	44%
0 - discharged with a non-COPD diagnosis	45	56%



## Name entity recognition (NER) tool:

- Named entity recognition tool: Clinical Language Annotation, Modeling and Processing (CLAMP).
- A general-purpose clinical NLP system built on proven methods:
- NLP tasks:
  - Named entity recognition: Rank #2 in 2009 i2b2 challenge, Rank#2 in 2010 i2b2 challenge
  - UMLS encoding: Rank #1 in 2014 SemEval challenge
  - Relation Extraction: Rank #1 in 2012 i2b2 Temporal challenge , Rank #1 in 2015 SemEval Disease-modifier challenge
- An IDE (integrated development environment) for building customized clinical NLP pipelines via GUI. It is a component-based system.



## Name entity recognition (NER) tool:

- These components: rule-based and deep-learning based
- Extracts the entities, extract relation between entities and assign a semantic tag to extracted entities (20 categories such as problem, test, drug, dosage).

The screenshot shows the OpenNLP GUI with the following components and pipeline configuration:

**Component List:**

- ML\_components (NLP Components)
  - NLP\_components
    - Sentence\_detector
    - Tokenizer
    - POS\_tagger
    - Chunker
    - Section\_identifier
    - Named\_entity\_recogizer
    - Assertion\_classifier
    - UMLS\_encoder
    - Relation\_extractor
    - Ruta\_rule\_engine

**Corpus Management:**

- Cancer
  - corpus
    - test
    - train
    - models

**Pipeline Configuration (lab-attribute.pipeline):**

Name	Component	Description
DF_Clamp_sentence_detector	Sentence detector	Rule based sentence detector
DF_Clamp_tokenizer	Tokenizer	Rule based tokenizer
DF_Dictionary_based_secti...	Section identifier	Dictionary based section header identifier
DF_OpenNLP_POS_tagger	POS tagger	OpenNLP based pos tagger
DF_Dictionary_lookup	Named entity recogizer	dictionary lookup algorithm
DF_CRF_based_named_ent...	Named entity recognizer	Name entity recognition using CRF
DF_Regular_expression_NER	Named entity recognizer	regular expression named entity recognition
DF_Ruta_script_file	Ruta rule engine	Ruta script
DF_Relation_connector_aft...	User Defined Components	set relation name after ruta script;

**DESCRIPTION:** Labtest-Value extraction  
**INPUT:** Encounter Notes/Discharge Summaries  
**OUTPUT:** Lab-Value pairs  
**CATEGORY:** General



## Name entity recognition (NER) tool:

- Leveraged rule-based and deep-learning based components of the CLAMP-GUI to extract the entities and assign them semantic tags.
- Updated its dictionary by manually checking the provider notes to find any verbiage for developing different pipelines.

The screenshot displays the CLAMP GUI interface. On the left, a 'Component' tree shows the 'Named\_entity\_recogizer' folder expanded, with an arrow pointing to the 'defaultDict.txt.zip' file. The main window shows a medical text document with various entities highlighted in yellow and orange, such as 'COPD', 'DM2', 'HTN', 'PE', 'dermatomyositis', 'synd', 'Complications', 'None', 'History', '65 yo', 'female', 'h/o', 'COPD', 'DM2', 'dermatomyositis', 'c/o', '1 month of increased cough and SOB', 'Also w/ chest/abd pain with cough', 'No f/c/purulent sputum', 'Was seen x several at outside ED's and given abx w/o relief', 'Sen at IMA on day of admit and sent to ED for admission', 'Admitting Examination: . Hospital Course and Results: 1.', 'COPD exac--Pt started on usual meds incl steroids, alb/atrvrent nebs and CTX/azithro', 'CXR with no acute changes--> likely bronchitis so CTX d/c'd', 'Pt improved on tx, to f/u with PMD and Pulm. 2. DM2--FS very'. Below the text, a 'Console' window shows the following output:

```
INFO: filename=[C:\Users\shahmf01\Downloads\ClampWin_1.6.6\workspace\MyPipeline\discharge_diagnosis\Data\Input\56338.txt]
INFO: filename=[C:\Users\shahmf01\Downloads\ClampWin_1.6.6\workspace\MyPipeline\discharge_diagnosis\Data\Input\56427.txt]
INFO: filename=[C:\Users\shahmf01\Downloads\ClampWin_1.6.6\workspace\MyPipeline\discharge_diagnosis\Data\Input\94392.txt]
```

Screenshot of CLAMP GUI output



## Feature Extraction

- Build different hybrid (machine learning + rules) pipelines for extracting:
  - Mentions of SDH (smoking, alcohol and drug status) from provider notes.
  - Very early vital signs from provider notes.
  - Very first symptoms documented in triage notes
- In addition to CLAMP defined output (offset, semantic tag, CUIs, negation and extracted entity), new tags will be added.

Screenshot of CLAMP  
Text output

Start	End	Semantic	CUI	Assertion	Entity
224	232	drug	C0028040, RxNorm=[7407]	present	nicotine
475	479	subject	null	null	wife
494	497	problem	null	present	ill
503	516	problem	null	absent	throat cancer



## Feature Extraction – SDH Pipeline Development

- Smoking status (Accuracy: 83%): developed a dictionary for new four tags as : “former smoker”, “current smoker”, “never smoked,” and “undetermined”. As a feature, we put the last 3 groups as patients that are not current smokers.
- Alcohol (Accuracy: 74%)and Drug status (Accuracy: 79%) : patients were categorized into three groups: “yes” (indicating that the patient consumes alcohol or has a drug misuse problem), “no” indicating the absence of these problems, and “undetermined” shows that the note doesn’t contain any mention for these conditions.

1	She has a prior history of <b>FormerSmoker</b> smoking although not currently .
2	He denies any <b>NeverSmoked</b> tobacco use .
3	He does have a history of <b>FormerSmoker</b> tobacco use and quit in 2006 .
4	SUBSTANCE USE : She has a past history of <b>FormerSmoker</b> tobacco use , but quit at the time of her myocardial infarction .
5	<b>CurrentSmoker</b> Smoker 1ppd x 15 years up to stemi 8/07



## Feature Extraction

- Separate pipeline to extract the vital signs from first provider/progress note.
- Vital signs: Blood pressure (BP), Pulse rate, respiratory rate (RR) and Spo2 value
- Extracted systolic and diastolic pressure values from BP value and added them to the feature set.

ID	Systolic Pressure	Diastolic Pressure	Pulse rate	Spo2	RR	Alcohol	Drug	Smoke
1	138	75	88	96	12	1	1	1
2	98	58	90	97	18	0	0	0
...	...	...	...	...	...	...	...	...
80	116	79	103	97	81	0	1	1

Note's extracted features



## Results & Discussion

- The most frequent reasons for the patient visit in both classes were shortness of breath (SOB) and COPD.
- Cough and wheezing were the next most common symptoms among the patients diagnosed with COPD exacerbation (class 1),
- For the patients in class 0, chest pain and cough were the next frequent symptoms.

Class 1			Class 0		
CUI	Entity	n	CUI	Entity	n
C0010399	SOB (shortness of breath)	21	C0010399	SOB (shortness of breath)	24
C0740304	COPD	13	C0740304	COPD	3
C0010200	Cough	5	C0008031	CP (chest pain)	3
C0043144	Wheezing	3	C0010200	Cough	2
C0004096	Asthma	3	C0020538	Hypertension	2
C0018802	CHF (congestive heart failure)	3	C0553668	labored breathing	2
C0553668	Labored breathing	2	C0038454	CVA (cerebrovascular accident)	1
C0747719	Poor appetite	2	C0853946	Worsening chest pain	1



## Results & Discussion

- Higher percentage of current smokers among the patients diagnosed with COPD exacerbation which is well aligned with typical COPD presentation.

Smoking	CI	CO	Alcohol	CI	CO	Drug	CI	CO
Former smoker	60%	60%	Yes	9%	37%	Yes	24%	46%
Current smoker	30%	23%	No	76%	57%	No	42%	46%
Never smoked	3%	6%	Undetermined	15%	6%	Undetermined	34%	8%
Undetermined	7%	11%						



## Discussion & Results

- Developed four predictive models based on naïve bayes, logistic regression, random forest, and SVM machine learning methods. F1-score and accuracy were used as the evaluation metrics.
- SVM resulted in the highest accuracy and F1-score (81% and 80%, respectively).

Models	F1-Score	Accuracy
Naïve Bayes	0.46	0.47
Logistic Regression	0.52	0.48
Random Forest	0.74	0.76
Support Vector Machine (SVM)	0.80	0.81

# MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA



---

Thank You!