Identifying Patients with Pneumonia from Free-Text Intensive Care Unit Reports

Abstract

Clinical research studying critical illness phenotypes relies on the identification of clinical syndromes defined by consensus definitions. Pneumonia is a prime example. Historically, identifying pneumonia has required manual chart review, which is a time and resource intensive process. The overall research goal of our work is to develop automated approaches that accurately identify critical illness phenotypes. In this paper, we describe our approach to the identification of pneumonia from electronic medical records, present our preliminary results, and describe future steps.

1. Introduction

Identification of complex clinical phenotypes among critically ill patients is a major challenge in clinical research. While large administrative datasets of Intensive Care Unit (ICU) patients exist, they lack the granular data necessary to accurately identify complex phenotypes and determine the relative timing of events during the course of critical illness. With the introduction of comprehensive electronic medical records (EMRs), all aspects of ICU care can now be captured in both structured and free-text format. The existence of such data provides an opportunity to identify critical illness phenotypes and facilitate clinical and translational studies of large cohorts of critically ill patients, a task that would not be feasible using traditional screening/manual chart abstraction methods. Our main research goal is to build automated tools to identify critical illness phenotypes and model their progression from ICU data. To accomplish this, we chose *pneumonia* as our first critical illness phenotype and conducted preliminary experiments to explore the problem space.

Current approaches use manual chart abstraction or bedside data acquisition to identify cases of pneumonia. This is labor intensive and involves many subjective assessments. In this paper, we will focus on identification of pneumonia based on the information available in various different types of reports created during the patient's ICU stay.

Under review by the International Conference on Machine Learning (ICML). Do not distribute.

2. Related Work

Several studies have demonstrated the value of Natural Language Processing (NLP) for a variety of health care applications (Demner-Fushman et al., 2009). One of those applications is infectious disease surveillance. Pneumonia surveillance is resource intensive. Within this domain, extraction of different types of pneumonia has been widely studied by various researchers. As one of the earliest examples, Fiszman et al. tested an NLP tool called SymText to identify acute bacterial pneumonia related concepts in chest x-ray reports and compared its performance against human annotation (Fiszman et al., 1999). Their results indicated that the performance of SymText was similar to that of the physician. The same research group used the concepts identified by SymText as features for automatically identifying chest x-ray reports that supported pneumonia (Chapman & Haug, 1999; Fizsman et al., 2000; Chapman et al., 2001; Aronsky et al., 2001). In their experiments, they compared the pneumonia classification performance of two machine learning algorithms (Decision Trees and Bayes Networks) and two rule-based approaches (simple keyword search and expert crafted rules). Their results showed that Bayesian networks perform as well as expert constructed systems and manual annotations performed by a physician and a lay person.

Another group of researchers investigated the feasibility of using NLP approaches in identifying healthcare associated pneumonia in neonates from chest x-ray reports (Mendonca et al., 2005; Haas et al., 2005). Their NLP approach involved two components: the MedLEE NLP system and rules that access the MedLEE output. The rules were manually constructed by a medical expert to identify chest x-ray reports indicating presence of pneumonia.

Elkin et al. also applied NLP approaches to identify pneumonia cases in free-text radiology reports (Elkin et al., 2008). Their system encoded the radiology reports with SNOMED CT Ontology and subsequently applied a set of manually constructed rules to the SNOMED CT annotations to identify the radiological findings and diagnoses related to pneumonia.

The studies outlined above have focused primarily on identification of pneumonia cases from radiologist chest x-ray reports. While radiologic changes within the lung are a necessary condition for diagnosis of pneumonia, there exists data within other domains such as the disease presentation narrative, physiologic measures, and laboratory abnormalities that could add significant accuracy and depth to the identification of pneumonia cases (Lutfiyya et al., 2006; Mandell et al., 2007). Because chest x-ray abnormalities comprise only part of the pneumonia definition, any system aimed at pneumonia identification which incorporates only chest x-ray information will lead to significant phenotypic misclassification. Thus, there remains an unmet need to accurately capture the clinical components of the pneumonia phenotype.

Physician daily notes are a potentially rich source of clinical information indicating the presence of phenotypes like pneumonia. In contrast to the narrow scope of information provided by radiology reports, physician daily notes include text detailing patient narrative, physiologic, imaging, and laboratory data, and, finally, the physician's interpretation of these data. We hypothesized that by using physician notes such as admit notes, ICU progress notes, and discharge summaries, automated approaches that incorporate NLP and machine learning can accurately identify pneumonia in ICU settings.

3. Methods

The overall architecture of our text processing approach for pneumonia extraction can be found in Figure 1. In the following sections, we will explain the main steps of the text processing approach in detail.

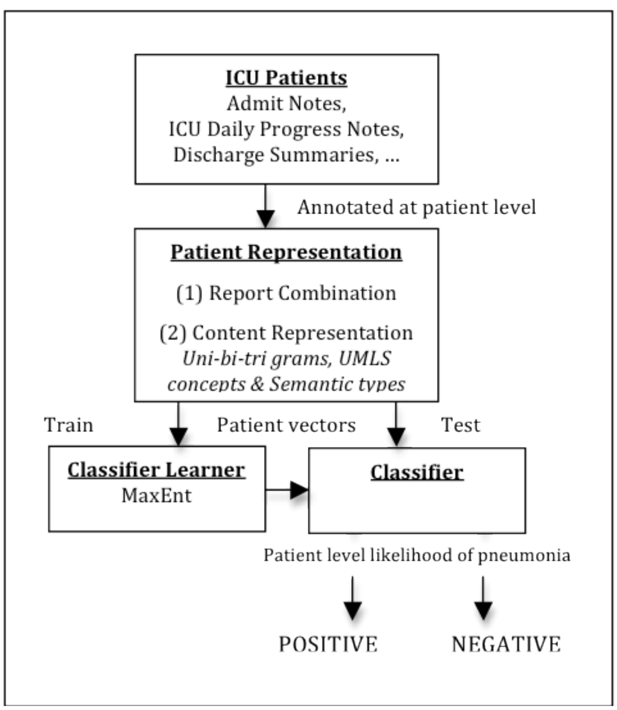

Figure 1. Overall system architecture of pneumonia classifier.

3.1 Data Set

The dataset was composed of 426 patients. The annotations used for this study were generated for another ongoing study of ICU subjects which has been described previously (reference withheld for review). An annotator with 6 years of experience as a research study nurse manually classified the positive and negative cases for pneumonia present at the time of ICU admission or within the first 48 hours of ICU stay (66 cases positive for pneumonia and 360 cases negative for pneumonia). The annotation was per-patient, and the annotator had access to the same set of text reports used for our NLP studies. However, the annotator did not perform sentence-level or report-level annotation for this corpus provided. Because annotation related to the presence or absence of pneumonia was limited to the period within 48 hours of ICU admission in this dataset and most patients were admitted to the ICU within 24 hours hospital admission (see Figure 2), we have targeted our pneumonia classifier to identify pneumonia occurring within 72 hours of hospital admission.

Pneumonia is defined as a lower respiratory tract infection which is associated with symptoms of acute infection with a new infiltrate on chest radiograph. Table 1 provides a summary of the characteristics of pneumonia.

Table 1. Characteristics of Pneumonia

Causes	
Bacteria: - *H. influenza* - *Strep pneumonia* - *Staph aureus* - Legionella species - Chlamydia species - *Pseudomonas aeruginosa*	Viruses: - Influenza - Parainfluenza Fungi: - Blastomycosis - Coccidiomycosis - Histoplasmosis
Clinical Signs and Symptoms	
Fever	Sputum production
Cough	Shortness of breath
Chest Pain	Malaise, fatigue
Abnormal white blood cell count	Muscle pains
Risk Factors	
Age > 65	
Immunosupression	
Recent antibiotic use	
Comorbid illnesses: HIV, Asthma, COPD, Renal Failure, Congestive Heath Failure, Diabetes, Liver Disease, Cancer, Stroke	

Pneumonia can be classified further based on the context in which it occurs. Community acquired pneumonia (CAP) refers to pneumonia that occurs outside of the hospital setting whereas hospital acquired pneumonia

(HAP) refers to pneumonia which occurs after admission to the hospital. Because subjects in this dataset were admitted to the ICU from the emergency department as well as from other hospitals, cases of pneumonia included both CAP and HAP.

Our dataset includes a total of 5313 reports from eight report types (admit note, ICU daily progress note, acute care daily progress note, transfer/transition note, transfer summary, cardiology daily progress note, and discharge summary) for 426 patients. The total number of reports per person ranged widely (median=8, interquartile range = 5-13, minimum =1, maximum=198). This is due to the high variability in the ICU length of stay.

The distribution among the eight different report types is presented in Table 2. The first column of the table gives the number of reports for each report type and the second column gives the number of distinct patients who had the report type in the dataset. As can be seen from the table, not all patients have all types of reports. As an example, only 280 (65% = 280/426) patients had admit notes; the remaining 146 patients had been transferred to the ICU from other medical units and therefore had no admit note. There were 350 (82% = 350/426) patients with discharge summaries. Of note, only a subset of 236 patients had both admit notes and discharge summaries (55% = 236/426).

Table 2. Report Statistics. Report Count: The frequency of report types, Patient Count: The number of distinct patients who had the report type.

Report Type	Report count	Patient Count
Admit Notes	481	280
ICU Daily Progress Note	2526	388
Acute Care Daily Progress Note	1357	203
Interim Summary	164	115
Transfer/Transition Note	243	175
Transfer Summary	18	18
Cardiology Daily Progress Note	133	17
Discharge Summary	391	350

The distribution of key note types by day after admission is shown in Figure 2. These histograms demonstrate that of the admit notes collected in this corpus, over 75% derive from the first day hospital admission. Furthermore, ICU progress notes were also consistently represented throughout the first 96 hours of admission. These data show that admit notes will be largely indicative of the pre-hospital and early hospital stay while ICU progress notes will be the predominant daily text source following the day of admit. Discharge notes largely arose >96 hours after admission (not shown).

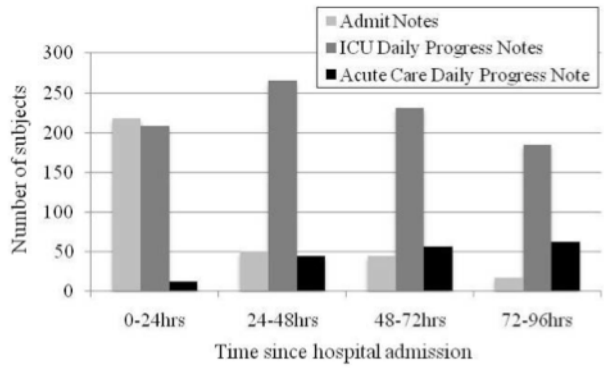

Figure 2. The distribution of report types in the first 96 hours of hospital visit.

3.2 Patient Representation

Representing the information available in the free-text reports is the most critical step in identifying patients with pneumonia. In our representation, we created one feature vector for each patient. There were two dimensions for experimentation: (1) finding the best combination of reports to be included when creating feature vectors, and (2) testing different text representation approaches to achieve the best classification performance.

3.2.1 Report Combination

As noted above, the presence/absence of pneumonia was manually annotated at the patient level, without a determination of which report(s) contributed to the phenotype determination. To understand how the contents of different types of reports contribute to the automated decision making, we tested different combinations of report types. Because the annotations for pneumonia in this dataset correspond to pneumonia present within the first 48 hours of ICU admission, we expect the admit notes to be particularly informative. Discharge summaries are expected to be informative because they contain the synthesis of numerous diagnostic studies and assessments conducted throughout the hospital course; however, they have the potential to introduce false positives due to pneumonia that occurred later in the hospital course. The following combinations were tested:

1. Only Admit Note: The patient vector was created only from the content of the admit notes of the patient.
2. Only Discharge Summary: The patient vector was created only from the content of the discharge summary.
3. Only Admit Note and Discharge Summary: The patient vector was created from the content of the admit note and discharge summary. Features from admit notes were separated from features from discharge summaries with labels indicating their

source. The features extracted from admit notes and discharge summaries had label formats *AdmitNote_$featureType_$featureName* and *DischargeSummary_$featureType$featureName*.

4. All Reports Combined: The patient vector was created from all the report types listed in Table 2 were merged under one document. All the features in the patient vector had the same label format *MergedNotes_$featureType_$featureName*.

Another important characteristic of our dataset is the availability of a timestamp on the creation of each report type, as well as a timestamp for the hospital admission and discharge. The majority of the patients had a series of ICU daily progress notes and acute care daily progress notes. In addition, transfer notes and inpatient notes from other departments (e.g. cardiology inpatient reports) describe the patient's status at the creation time of the report. By calculating the difference between the patient's hospital admit timestamps and report creation timestamps, we calculated the relative time each report was created during a patient's ICU stay (e.g., day 1, day 2) and defined the following three report combination alternatives.

5. Admit Note, Discharge Summary, and Other Days Separated: in this setting, we have many separate report categories, the admit note, the discharge note, and the collection of reports for each specific day of the ICU stay. Features were labeled as *AdmitNote_$featureType_$featureName*, *DischargeSummary_$featureType$featureName*, and *OtherNotes_$day_$featureType_$featureName*.

6. Admit Note, Discharge Summary, and Others within initial 72 hours of hospital admission: in this setting, we have 3 report categories, the admit note, the discharge note and the collection of reports whose timestamp is within the first 72 hours of hospital admission. This time frame is consistent with our annotations indicting the presence of pneumonia within the first 48 hours of ICU stay given that most subjects were admitted to the ICU within 24 hours of hospital admission. Features in the patient vector were labeled as *AdmitNote_$featureType_$featureName*, *DischargeSummary_$featureType$featureName*, and *OtherNotesPRE72_$featureType_$featureName*.

7. Admit Note, Discharge Summary, Others within initial 72 hours, and Others Post 72 hours: in this setting, we have 4 report categories, the admit note, the discharge note, the collection of reports whose timestamp is with the first 72 hours of admission as well as the collection of reports for the subsequent hospital stay. Features were labeled as *AdmitNote_$featureType_$featureName*, *DischargeSummary_$featureType$featureName*, *OtherNotesPRE72_$featureType_$featureName*, and *OtherNotesPOST72_$featureType_$featureName*.

3.2.2 Content Representation

Content representation has a direct effect on the overall classification performance and we used the following feature types in our representation.

Baseline Features

Information retrieval research suggests that words (uni-grams) work well as representation units for retrieving documents (Lewis, 1992). We used the uni-grams as the feature baseline in our experiments. We used a list of common English stopwords[1] to filter the stopwords from unigrams.

N-Gram Features

We used word bi-gram and tri-gram features to capture interesting multi-word features.

Knowledge-based Features

Representing the content with a bag-of-words approach has two challenges. First, the percentage of multi-word phrases, such as *community acquired pneumonia*, in medical vocabulary is very high. This high prevalence of phrases represents a problem for classification. For example, the meaning of *community acquired pneumonia* is very different from that of *community* alone. Second, synonymy is a very common characteristic among the medical phrases. For example, clinicians use *liver* and *hepatic* interchangeably even in the same reports. If not grouped explicitly, synonymous words or phrases are represented as different features in the feature vector, which has two major drawbacks. The first drawback is that the increase in the dimensionality of feature space is known to have a negative effect on classification performance. The second drawback is that information is lost due to feature splits. Instead of having a stronger feature, the representation has multiple relatively weaker features that are synonyms of each other.

To identify biomedical phrases, our system uses a knowledge-based NLP tool to process each report sentence and identify the domain specific terms. The biomedical domain already has a large publicly available knowledge base called the Unified Medical Language System (UMLS)[2]. In the latest version of UMLS, there are over 2.3 million biomedical concepts as well as over 8.5 million concept names. To identify the biomedical phrases, we used MetaMap, a tool created by NLM that maps the strings in free text to biomedical concepts in the UMLS[3]. MetaMap uses the UMLS to find the closest

[1] English Stopword List. Available at: ftp://ftp.cs.cornell.edu/pub/smart/english.stop

[2] Unified Medical Language System Fact Sheet. Available at: http://www.nlm.nih.gov/pubs/factsheets/umls.html

[3] Metamap (MMTx). Available at:http://mmtx.nlm.nih.gov/

matching known concept to each identified phrase in the free text. Our system sends each sentence to MetaMap and uses the Concept Unique Identifier (CUI) of the identified UMLS concepts to group the synonymous concepts. We used the identified CUIs as binary features in our representation.

Another knowledge source available in UMLS is the Semantic Network, which is a directed graph composed of 135 categories called *semantic types* and 49 different relations defined between the semantic types. Each medical concept in UMLS is mapped to at least one semantic type. For example, the medical concept *pneumonia* has a semantic type "*disease or syndrome"*, and the concept *magnesium* has two semantic types of "*biologically active substance"* and "*element, ion, or isotope"*. We used the semantic types of the extracted UMLS concepts from medical records to represent the content of medical records.

3.3 Pneumonia Classifier

After representing the content of the available reports with baseline, n-gram, and knowledge-based features, we trained classifiers to identify the patients with pneumonia within the first 48 hours of ICU stay. For our classification task, we picked the Maximum Entropy (MaxEnt) algorithm due to its good performance in text classification tasks (Nigam et al., 1999). In our experiments, we used the MaxEnt implementation in a machine learning package called Mallet[4].

4. Results

In our experiments, we used the patient level pneumonia annotations described in Section 3.1. as the gold standard to train classifiers and to test their performance.

4.1 Metrics

We evaluated the classification performance by using precision, recall, F1, specificity (proportion of negatives which are correctly identified), and accuracy performance metrics. Because there was a limited number of positive cases in our dataset, we decided to use 5-fold cross validation to measure the overall classification performance.

4.2 Classification Performance

We measured the MaxEnt classification performance for the different types of report combinations and feature types described in the previous section.

4.2.1 Report Combination

To understand the contribution of different report types on the classification performance, we compared different combinations of reports described in section 3.2.1, using only the unigram baseline features. Table 3 includes the performance values for the cases where patient vectors were generated (1) *only from admit notes*, (2) *only from discharge summaries*, (3) *only from admit notes and discharge summaries*, and (4) *all reports merged without distinguishing report type*. We ran these experiments for the 236 patients who had both admit notes and discharge summaries (44 cases positive for pneumonia and 192 cases negative for pneumonia). We represented the content of the reports with baseline unigram features.

As can be seen in Table 3, the performance of the classifier trained solely on admit notes is higher than that trained on discharge summaries only. However, we see that the highest performance is for the classifier trained on all of the reports without distinguishing report type.

In our second comparison, we investigated the effect of introducing report timestamp information in the feature representation. We ran this experiment using all 426 patients in our dataset. We compared (1) *all reports merged without distinguishing report type*, (2) *admit note, discharge summary, and others days separated*, (3) *admit note, discharge summary, and others within initial 72 hours of hospital admission*, and (4) *admit note, discharge summary, others within initial 72 hours, and others post 72 hours.* Table 4 includes the performance values. As can be seen from the table, we receive the lowest F1 performance with (3) *admit note, discharge summary, and others within initial 72 hours of hospital admission.* Adding post 72 hours reports (4) decreases precision but increases recall due to an increased number of false positives, as expected. Including all reports separated by days (2), increases both precision and recall when compared to (3) and (4). However, the best performance results are seen with all note types merged (1).

4.2.2 Content Representation

To determine the effect of different feature types described in Section 3.2.2, we conducted various experiments with different feature combinations. As can be seen in Table 5, adding bigrams (2) and trigrams (3) to unigrams (1), does not affect the precision but decreases the recall. Although the size of the feature space with UMLS concepts is much smaller than that of unigrams (Table 6), UMLS concepts (4) and unigrams (1) perform similarly both in terms of recall and precision.

When UMLS concepts and unigrams are combined (6), there is no significant performance change. When semantic types are added to the UMLS concepts (5), the recall stays the same, but precision decreases. The best precision is achieved when unigrams are combined with UMLS concepts (6) and best recall is achieved when unigrams, UMLS concepts and semantic types are combined (7).

[4] Mallet. Available at: http://mallet.cs.umass.edu

Table 3. Performance evaluation of different report combinations based on report type with baseline uni-gram features. The experiments were run for the 236 patients with both admit notes and discharge summaries. TP: True positive, TN: True negative, FP: False positive, FN: False negative, PRE: precision, REC: recall, F1: F1-Score, SPE: specificity, ACC: Accuracy. The lowest value for FP and FN is in boldface. The highest value for the remaining columns is in boldface.

REPORT COMBINATION	TP	TN	FP	FN	PRE	REC	F1	SPE	ACC
(1) ADMIT NOTE ONLY	13	182	10	31	56.5	29.5	38.8	94.8	82.6
(2) DISCHARGE NOTE ONLY	7	175	14	37	33.3	15.9	21.5	92.6	78.1
(3) ADMIT + DISCHARGE NOTE ONLY	8	174	18	36	30.8	18.2	22.9	90.6	77.1
(4) ALL NOTE TYPES MERGED	**17**	**186**	**6**	**27**	**73.9**	**38.6**	**50.7**	**96.9**	**86.0**

Table 4. Performance evaluation of different report combinations based on timestamps with baseline uni-gram features. The experiments were run for all 426 patients. TP: True positive, TN: True negative, FP: False positive, FN: False negative, PRE: precision, REC: recall, F1: F1-Score, SPE: specificity, ACC: Accuracy. The lowest value for FP and FN is in boldface. The highest value for the remaining columns is in boldface.

REPORT COMBINATION	TP	TN	FP	FN	PRE	REC	F1	SPE	ACC
(1)ALL NOTE TYPES MERGED	**28**	**340**	**20**	**38**	**58.3**	**42.4**	**49.1**	**94.4**	**86.4**
(2) ADMIT + DISCHARGE NOTE + OTHERS DAY SEPARATED	23	336	24	43	48.9	34.8	40.7	93.3	84.3
(3) ADMIT + DISCHARGE NOTE + OTHERS PRE 72 HOURS	19	338	21	47	47.5	28.8	35.8	94.2	84.0
(4) ADMIT + DISCHARGE NOTE + OTHER PRE 72 HOURS + OTHERS POST 72 HOURS	22	331	29	44	43.1	33.3	37.6	91.9	82.9

Table 5. Performance evaluation of different feature combinations with all note types combined. The experiments were run for all 426 patients. TP: True positive, TN: True negative, FP: False positive, FN: False negative, PRE: precision, REC: recall, F1: F1-Score, SPE: specificity, ACC: Accuracy. The lowest value for FP and FN is in boldface. The highest value for the remaining columns is in boldface.

CONTENT REPRESENTATION	TP	TN	FP	FN	PRE	REC	F1	SPE	ACC
(1) UNIGRAMS	28	340	20	38	58.3	42.4	**49.1**	94.4	**86.4**
(2) UNI + BIGRAMS	23	342	18	43	56.1	34.9	43.0	95.0	85.7
(3) UNI + BI + TRIGRAMS	21	**345**	**15**	45	58.3	31.8	41.2	**95.8**	85.9
(4) UMLS CONCEPTS	28	339	21	38	57.1	42.4	48.7	94.2	86.2
(5) UMLS CONCEPTS + SEMANTIC TYPES	28	330	30	38	48.3	42.4	45.2	91.7	84.0
(6) UNI-GRAMS + UMLS CONCEPTS	27	341	19	39	**58.7**	40.9	48.2	94.7	**86.4**
(7) UNI -GRAMS + UMLS CONCEPTS + SEMANTIC TYPES	**31**	329	31	**35**	50.0	**47.0**	48.4	91.4	84.5
(8) UNI+BI +TRIGRAMS + UMLS CONCEPTS + SEMANTIC TYPES	25	327	33	41	43.1	37.9	40.3	90.8	82.6

Table 6. Feature set sizes (all note types merged) for 426 patients.

FEATURE TYPE	# OF DISTINCT FEATURES
UNIGRAM	30751
BIGRAMS	361357
TRIGRAMS	824748
UMLS CONCEPTS	19546
UMLS SEMANTIC TYPES	125

Table 7 includes the top 25 ranked features from different feature types and their combinations. As can be seen from the table, many of the features identified by the classifier are closely linked to the known causes (e.g. *influenza*) and clinical signs & symptoms (e.g. *cough, sputum*) of pneumonia listed in Table 1. In addition, medications including antibiotics commonly used for the treatment of pneumonia (e.g. *levofloxacin, vancomycin*) were included among the most predictive features identified by the classifier. Interestingly, several features related to alteration in level of alertness were also included (e.g.

anoxic, *encephalopathy* and *AMS* [shorthand for altered mental status]). While these terms do not relate directly to the diagnostic criteria for pneumonia, they may well indicate latent risk factors for pneumonia related to a decreased level of consciousness and reduced ability to protect the respiratory tract from contamination from the upper airway.

4.3 Discussion

Our best system achieved a F1-score of 49.1% for the 426-patient dataset. While the result is encouraging, there is still much room for improvement. Two main factors contribute to the errors made by our current system: the limitations of the data set and the system design.

There are several important limitations to our current dataset. First, it is not a complete set of reports for all patients (e.g., any notes entered prior to the day of ICU admission were not captured). In our dataset, 146 (35%=146/426) patients did not have admit notes and 76 (18%=76/426), patients did not have discharge summaries and 77 (18%=77/426) patients did not have any reports generated in the first 24 hours of hospital stay. Second, the pneumonia annotation in this dataset was created for a different purpose, where the annotator (a medical expert) was asked to determine whether a patient had pneumonia within 48 hours of admission to the ICU. In contrast, our system used notes within 72 hours of hospital admission as a proxy for this period. Therefore, there is a potential mismatch between the annotation and our task for the minority of cases in which the gap between hospital admission and ICU admission was greater than 24 hours. Third, the data set is relatively small with a limited number of positive pneumonia cases.

Another source of the system errors is due to the limitations of our current system, which relies on features available from shallow processing of the text. The detection of a phenotype such as pneumonia often requires a deeper understanding of the reports. For instance, as shown in Table 7, the word unigram *pneumonia* (PNA for short) or its corresponding UMLS concept is a strong feature that indicates that the patient

Table 7. Top 25 ranked features for MaxEnt models with different feature types and combinations.

Rank	UniGram	Uni-Bi-TriGrams	UMLS Concepts	Uni-Bi-TriGrams + UMLS Concepts + Semantic Types
1	pneumonia	pneumonia (uni)	C0032285-Pneumonia	C0430400-Laboratory culture
2	sputum	sputum (uni)	C0430400-Laboratory culture	T074-Medical Device
3	aspiration	aspiration (uni)	C0580264-H1N1	C0032285-Pneumonia
4	cx	influenza (uni)	C0035410-Rhabdomyolysis	T031- Body Substance
5	influenza	continue (uni)	C0021400-Influenza	C0524425 - Endovascular
6	day	day (uni)	C0027552-Needs	T184-Sign and Symptom
7	continue	perirectal (uni)	C0038056-Sputum	C0580264-H1N1
8	perirectal	cough (uni)	C0032290-Aspiration Pneumonia	T116-Amino Acid, Peptide, or Protein
9	cough	oseltamivir (uni)	C0332148-Probable diagnosis	T083- Geographic Area
10	ddavp	mg_po (bi)	C0021403-Influenza virus vaccine	pneumonia (uni)
11	lisinopril	lisinopril (bi)	C0948187-Tracheomalacia	C0038257-stent
12	gpc	aspiration_pneumonia (bi)	C0019134-Heparin	T028 - Gene or Genome
13	levofloxacin	metoprolol (uni)	C0038846-Supine Position	continue (uni)
14	shunt	requiring (uni)	C0003980-Asia	T195-Antibiotic
15	metronidazole	gpc (uni)	C0234422-Awake	T055- Individual Behavior
16	needed	tube (uni)	C0600500-Peptide Nucleic Acids	T197- Inorganic Chemical
17	anoxic	feeding tube (bi)	C0392747-changing	T129- Immunologic Factor
18	po	ddavp (uni)	C0699992-lasix	aspiration (uni)
19	water	needed (uni)	C0558288-as required	T185-Classification
20	porequiring	vancomycin (uni)	C1550291-Perirectal	C0038056-Sputum
21	ams	levofloxacin (uni)	C1547295-Acute	T022-Body System
22	metoprolol	shunt (uni)	C0005889-body fluids	C0027552-Needs
23	encephalopathy	anoxic (uni)	C0087111-Therapeutic procedure	C1550291-Perirectal
24	vancomycin	encephalopathy (uni)	C0000737-Abdominal Pain	C0231835-Tachypnea
25	cmt	as needed for (tri)	C1547229-Acute	T114-Nucleic Acid, Nucleoside, or Nucleotide

has PNA. But there are many contexts where the presence of the word PNA does not mean that the patient has PNA within the 48 hours of admission to ICU. Some examples are explicit or implicit negation (*The lab result is inconsistent with PNA*; *PNA is ruled out*), past history (*He had PNA two years ago*), family history (*His father had PNA*), possibility (*Action items: PNA or flu*), PNA appearing in reports after the 48-hour window, and so on. Another challenge is that sometimes the ICU reports mentioned the PNA-related symptoms or lab results without explicitly stating that the patient had PNA, so it is important for our system to identify these symptoms and lab results. However, the lab results and other structured data (e.g. temperature, blood pressure) are often not included in the ICU text report. While the objective of the current study was to explore the text reports using NLP and machine learning methods, our ultimate goal is to combine these notes with informative structured data and radiology reports. We hypothesize that this combined approach will be superior to using text reports or structured data alone.

5. Conclusion

In this paper, we described a text processing approach to identify the cases with pneumonia from the information available in eight different reports types generated in the ICU setting. We tested various different report combinations and feature types to increase the performance of our tools.

In this paper, we presented our preliminary results. Although our dataset had significant limitations, the results are encouraging. Our best performing classifier based on F1 produced 58.3% precision, 42.4% recall, 49.1% F1, 94.4% specificity, and 86.4% accuracy. As future work, we will focus on the following areas. First, we will create a new dataset with both patient level and report level phenotype annotation. Second, we will extend the current system to include features generated from a deeper processing of the free-text reports to discover information such as scope of negation and lab test results. Third, we will combine NLP processing of radiology reports and structured data elements (e.g. white blood count, temperature) available in EMR with the information extracted from free-text reports to improve the classification of pneumonia.

References

Aronsky D, Fiszman M, Chapman WW, Haug PJ. Combining decision support methodologies to diagnose pneumonia. *Proc AMIA Symp*, pp. 12-16, 2001.

Chapman WW, Haug PJ. Comparing expert systems for identifying chest x-ray reports that support pneumonia. *Proc AMIA Symp*, pp.216-220, 1999.

Chapman WW, Fiszman M, Chapman BE, Haug PJ. A comparison of classification algorithms to automatically identify chest x-ray reports that support pneumonia. *Journal of Biomedical Informatics*, 34: 4-14, 2001.

Demner-Fushman D, Chapman WW, McDonald C. What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*; 42(5): 760-772, 2009.

Elkin PL, Froehling D, Wahner-Roedler D, Trusko B, Welsh G, Ma H, Asatryan AX, Tokars JI, Rosenbloom ST, Brown SH. NLP-based identification of pneumonia cases from free-text radiological reports. *Proc AMIA Symp*, pp. 172-176, 2008.

Fiszman M, Chapman WW, Evans SR, Haug PJ. Automatic identification of pneumonia related concepts on chest x-ray reports. *Proc AMIA Symp.* pp. 67-71, 1999.

Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest x-ray reports. *Journal of the American Medical Informatics Association*, 7(6): 593-604, 2000.

Haas JP, Mendonca EA, Ross B, Friedman C, Larson E. Use of computerized surveillance to detect nosocomial pneumonia in neonatal intensive care unit patients. *Am J Infect Control*, 33(8): 439-443, 2005.

Lewis DD. An Evaluation of Phrasal and Clustered Representations on a Text Categorization Task. *Proc. of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 37-50. 1992.

Lutfiyya MN, Henley E, Chang LF, Reyburn SW. Diagnosis and treatment of community-acquired pneumonia. *Am Fam Physician*, 73(3): 442-50, 2006.

Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, Dowell SF, File TM Jr, Musher DM, Niederman MS, Torres A, Whitney CG. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis.*, 44 Suppl 2:S27-72, 2007.

Mendonca EA, Haas J, Shagina L, Larson E, Friedman C. Extracting information on pneumonia in infants using natural language processing of radiology reports. *Journal of Biomedical Informatics*, 38(4): 314:321, 2005.

Nigam K, Lafferty J, Mccallum A. Using maximum entropy for text classification. *Proc. IJCAI-99 Workshop on Machine Learning for Information Filtering*, pp. 61-67, 1999.