Supplementary Material A: Search Terms for Pneumonia

"%PNA%"
"%HCAP%"
"%bronchopneumonia%"
"%pleuropneumonia%"
"%legionellosis%"
"%legionnaire%"
"%pneumonitis%"
"%pneumonia%"
"%pneumoniae%"

"misspellings /
"%pneumonia%"
"%pneumo%"
"%pnumonia%"
"%pnemonia%"
"%pneuonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumoni%"
"%ppneumonia%"
"%pnneumonia%"
"%pneeumonia%"
"%pneuumonia%"
"%pneummonia%"
"%pneumomin%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumoni%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
"%pneumonia%"
Supplementary Material B: Pneumonia Annotation Guideline

Thank you for helping to annotate for this project! Our goal is to develop a natural language processing tool that can classify emergency department (ED) clinician-written documents on whether they contain a clinical assertion of the diagnosis of pneumonia in the ED.

The question we are asking is:

"Does the clinician assert the patient has pneumonia at the time of the documentation/encounter?"

We are not asking:

"Does the patient actually have evidence for pneumonia?" / "Do you think the patient actually has pneumonia?"

"Does the patient have a history of pneumonia?"

Terms used for pneumonia:

Pneumonia, pna, HCAP ("health care-associated pneumonia")
parapneumonic effusion = collection of fluid between chest wall
empyema = collection of pus

If a clinician includes "infection/empyema" or "parapneumonic effusion" in their diagnosis, they think the patient has pneumonia.

"Pneumonitis" is NOT necessarily pneumonia, unless the clinician specifically states that it is an "infectious pneumonitis." Noninfectious pneumonitis includes aspiration, postobstructive, or chemical pneumonitis that is not actually pneumonia.
We have identified four classes:

1. Clearly pneumonia—clinician is certain of the diagnosis of pneumonia and proceeds with treatment without mentioning any other possibilities.
2. Possible pneumonia-treated—clinician thinks the dx of pneumonia is possible or probable, and proceeds with treatment.
3. Possible pneumonia-not treated—clinician considers the dx of pneumonia but does not treat.
4. Clearly not pneumonia—clinician indicates a different diagnosis and does not list pneumonia as a possible diagnosis, or mentions that it was specifically excluded.

**General Principles**

Use clinician’s language (assessment/suspicion/assertion), rather than objective information (evidence).

The majority of the relevant text will be in the assessment/plan, or the impression, which is usually at the end of the note.

**Spans to Annotate**

*Annotate the entire assessment* if pneumonia is lacking (this is a clear negative).

*Annotate parts of the assessment* that contain the suspicion of pneumonia if it is possible or clearly positive.

*Annotate the plan* if it includes antibiotics that the clinician is using to cover patient for pneumonia mentioned in the assessment (this is a clear positive or possible-treated).

Annotate mentions of pneumonia as history in the history of present illness or past medical history (PMH, “recently hospitalized for pneumonia,” “thought he had pneumonia,” “was in overnight with diagnosis of pneumonia”) as clearly not pneumonia.

Annotate mentions of pneumonia from the same visit.

Annotate “pneumonia vaccine” as clearly not pneumonia.

Spans NOT to annotate:

- Do not annotate chief complaint or reported symptoms.
- Do not annotate chest imaging report results.
- Do not annotate physical exam findings.
- Do not annotate vitals/laboratories.
- Do not annotate medication lists that are not in the plan (i.e., the medications the patient was taking prior to the visit).

*Do not annotate information in any notes that are not physician notes.* Classify them as “not MD note,” and move on.

Do not annotate patient instructions/hypothetical (return to ER if...)

**Document Classification**

Select the last word of the entire document by double-clicking on it. Select “Document_Classification.” Click on attributes, and select the final classification from pull-down menu.

If the note is clearly not a physician note (e.g., it is an addendum showing XR result, or nursing not, or laboratory results, etc.), classify as “Not MD note” and do NOT annotate.

**Definitions**

1. “Not MD Note”

We aim to annotate only those documents that actually reflect the medical decision making of the physician/PA’s/NP’s that evaluated the patient. The typical structure of the document should be Chief complaint, HPI, medical history/medications, physical exam, laboratories/studies, and assessment plan. If the document does not contain a clinical assessment by a physician/PA/NP, DO NOT ANNOTATE and classify it as “Not MD Note.” Example: document written by a nurse (communication/triage note); document by a physician/nurse that does not contain an assessment/plan, such as “admission evaluation” form or addendum.

1. Clear Positives

1. Assessment includes pneumonia as leading diagnosis with no other possible diagnoses offered.

2. Treat resolving pneumonia as clear pneumonia—but if it says *resolved* pneumonia, this is a clear negative.

3. COPD and pneumonia, or COPD secondary pneumonia—COPD and pneumonia can occur together and should be designated as clear pneumonia. However, if it is COPD versus pneumonia, then this is not clear pneumonia and should be classified as a possible (-treated if given abx, not treated if not given abx).
2. Possible-treated

1. “Probable pneumonia,” “possible HCAP.”
2. “More likely pneumonia,” “concern for infection/empyema.”
3. Assessment includes pneumonia as leading diagnosis but offers additional possibilities.
4. Assessment includes pneumonia as secondary diagnosis and indicates that they are proceeding with pneumonia-specific treatment (antibiotics).
   a. **If you truly cannot determine whether antibiotics were administered, and you think the clinician suspected pneumonia, then indicate possible-treated.**
5. Clinician indicates that they are questioning a diagnosis of pneumonia made by another clinician, but they are proceeding with treatment.
6. Assessment includes pneumonia as a possible diagnosis and clinician indicates that he/she is proceeding with pneumonia-specific treatment.
   a. Pneumonia-specific treatment = antibiotics including:
      i. Ceftriaxone, Rocephin, Azithromycin, Vancomycin, piperacillin/tazobactam, levofloxacin, moxifloxacin, doxycycline.
   b. If the antibiotic is not specified (e.g., “Plan: antibiotics,” then assume those antibiotics are for pneumonia if it is listed as a possible diagnosis.

2. Possible-not treated

1. Assessment that includes pneumonia as a possible alternative diagnosis but clinician indicates that he/she is not proceeding with pneumonia-specific treatment.
2. Clinician indicates that they are questioning a diagnosis of pneumonia made by another clinician, and they are stopping treatment.
3. “Rule out (or r/o) pneumonia”—clinician thinks that pneumonia is possible, and indicates that he/she will proceed to diagnostic testing (chest X-ray or CT), but does not indicate that they are treating.
4. Clinician indicates that pneumonia is possible and considers evaluation (such as obtaining a cxr) but does not indicate that he/she will be treated for pneumonia.
1. Clear negatives: does the clinician think the patient clearly does NOT have pneumonia?
   1. Lacks mention of pneumonia in differential diagnosis or final assessment.
   2. Clinician states that pneumonia has been ruled out.
   3. Pneumonia mentioned in a history (“past medical history of pneumonia,” “recently hospitalized for pneumonia.”
   4. Mentions pneumonia in recent history or past medical history, but not in the assessment of the current presentation.
   5. Clinician mentions the possibility of pneumonia but states that this possibility was ruled out/excluded (“other possible causes, pneumonia, which is excluded by the chest X-ray”).
5. Inconsistencies/ambiguities
Often, the emergency department note contains a “medical decision-making/emergency department course” section, containing an initial differential diagnosis before all of the studies/work-up comes back. This working diagnosis can be inconsistent with the clinical final impression, which is typically at the end of the note. To make the document-level classification, if there is discordance, use the final impression rather than the medical decision-making assertions.

Often, the X-ray/radiographic imaging report is pasted into the note, and this often contains an “Impression” made by the radiologist (i.e., “Impression: no pneumonia” or “Impression: RLL infiltrate suspicious for pneumonia”). Be careful! This is NOT the same as the documenting physician’s impression/assessment. Do not annotate these mentions of pneumonia.

Scroll to the end of the note to find the physician’s assessment.

If the clinician seems to be inconsistent in his/her assessment, take the latest-appearing statement as their final clinical suspicion.

Annotate all statements, but classify the document based upon the final clinical suspicion.

Example: If the initial impression is possible pneumonia but the final impression is clearly not pneumonia (for example, if the clinician ruled pneumonia after physical exam and CXR), the document-level classification is clearly not pneumonia/not treated.

Supplementary Material C: Complete Contingency Tables

Table S1 Training Set

<table>
<thead>
<tr>
<th>NLP ICD9 and Human Classifications in 811 cases with documents containing search terms for pneumonia</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ED-ICD-9</td>
<td>Human</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.477</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>387</td>
<td>216</td>
<td>603</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>197</td>
<td>208</td>
<td>PPV</td>
</tr>
<tr>
<td>398</td>
<td>413</td>
<td>811</td>
<td>NPV</td>
<td>0.642</td>
</tr>
<tr>
<td>Hosp-ICD9 (N = 567)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>217</td>
<td>181</td>
<td>398</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>151</td>
<td>169</td>
<td>PPV</td>
</tr>
<tr>
<td>235</td>
<td>332</td>
<td>567</td>
<td>NPV</td>
<td>0.545</td>
</tr>
<tr>
<td>NLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.724</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>379</td>
<td>114</td>
<td>493</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>299</td>
<td>318</td>
<td>PPV</td>
</tr>
<tr>
<td>398</td>
<td>413</td>
<td>811</td>
<td>NPV</td>
<td>0.769</td>
</tr>
<tr>
<td>NLP + EDICD9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.879</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>368</td>
<td>50</td>
<td>418</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>363</td>
<td>393</td>
<td>PPV</td>
</tr>
<tr>
<td>398</td>
<td>413</td>
<td>811</td>
<td>NPV</td>
<td>0.880</td>
</tr>
</tbody>
</table>

Table S2 Validation Set

<table>
<thead>
<tr>
<th>NLP ICD9 and Human Classification of 100 randomly selected cases (needed to have an ED note–total N = 2,881,470)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ED-ICD-9</td>
<td>Human</td>
<td>Totals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.444</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>91</td>
<td>5</td>
<td>96</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>PPV</td>
</tr>
<tr>
<td>91</td>
<td>9</td>
<td>100</td>
<td>NPV</td>
<td>0.948</td>
</tr>
<tr>
<td>Hosp-ICD9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.667</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>90</td>
<td>3</td>
<td>93</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>PPV</td>
</tr>
<tr>
<td>90</td>
<td>9</td>
<td>100</td>
<td>NPV</td>
<td>0.968</td>
</tr>
<tr>
<td>Search term for pneumonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>78</td>
<td>0</td>
<td>78</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>9</td>
<td>22</td>
<td>PPV</td>
</tr>
<tr>
<td>91</td>
<td>9</td>
<td>100</td>
<td>NPV</td>
<td>1.000</td>
</tr>
<tr>
<td>NLP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.556</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>89</td>
<td>4</td>
<td>93</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>PPV</td>
</tr>
<tr>
<td>91</td>
<td>9</td>
<td>100</td>
<td>NPV</td>
<td>0.957</td>
</tr>
<tr>
<td>NLP plus ICD9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>Sensitivity</td>
<td>0.889</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>89</td>
<td>1</td>
<td>90</td>
<td>Spec</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>PPV</td>
</tr>
<tr>
<td>91</td>
<td>9</td>
<td>100</td>
<td>NPV</td>
<td>0.989</td>
</tr>
</tbody>
</table>