Project Proposal

Chromilo Amin | Friday, March 26, 2021 | Bertelsmann Technology Scholarship program

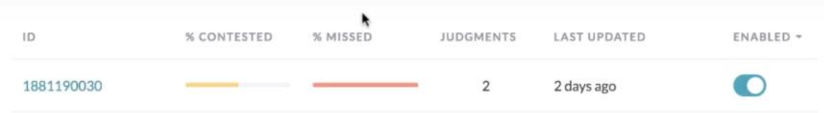
Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	The goal is to build a product that will help doctors quickly identify cases of pneumonia in children. We hope to train ML to correctly identify with high confidence level the presence of pneumonia from a set of images. The images will have a mix of healthy and abnormal lungs. In summary, • help flag serious cases, • quickly identify healthy cases, • and, generally, act as a diagnostic aid for doctors.
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels vs any other option?	Cloudiness_present – this is a yes-no selection Diapragm_present – this is a yes-no selection Heart_present - this is a yes-no selection Cloudy_types this is a checkbox selection asking annotator to indicate whether the cloudiness is a big patch or various small patches, and on which side of the lung Confidence_level – this is a number from 1-3 with 1 being not confidence, 2 as somewhat confident, and 3 the most confident in the answers. Confidence_detail – this is a text field that comes up if the annotator picked 2 for confidence level above. The pros for the above approach: 1- Simple yes/no questions avoids grey areas; 2- Unknown categories show up in confidence level, again kept simple with 1-3 only options with 2 being unsure. 3- Free text field gives annotators an area to voice their reason for being unsure on their answers. The cons for the above approach: 1- Being overly simple especially in confidence_level ranking could result in further followups to determine if they are more or less sure about how unsure they are, i.e. expanding to 1-5 or even 1-10 choices may be improve ranking. 2- Lack of free text fields for each data label could hinder question improvements and increases time to surveys at the end of the job.

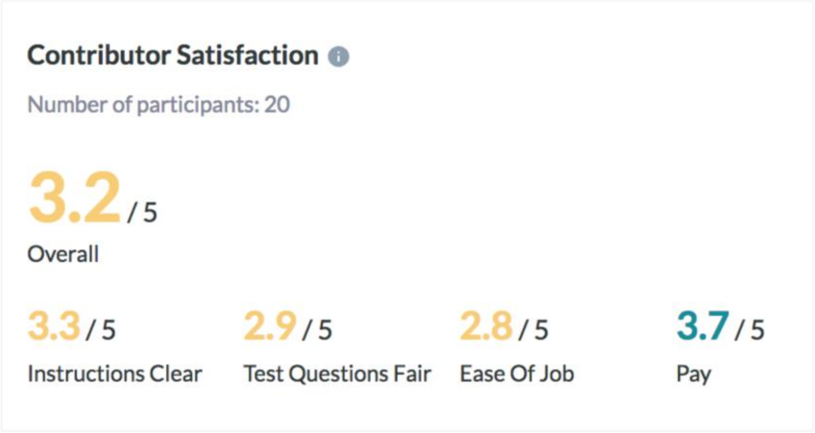
Test Questions & Quality Assurance

<table>
<tr>
<td><b>Number of Test Questions</b>

Considering the size of this dataset, how many test questions did you develop to prepare for launching a data annotation job?</td>
<td>I picked 3 images with abnormal readings and 2 with healthy readings for a total of 5-6 test question samples.</td>
</tr>
<tr>
<td><b>Improving a Test Question</b>

Given the following test question which almost 100% of annotators missed, statistics, what steps might you take to improve or redesign this question?</td>
<td>
<table>
<tr><th>ID</th><th>% CONTESTED</th><th>% MISSED</th><th>JUDGMENTS</th><th>LAST UPDATED</th><th>ENABLED</th></tr>
<tr><td>1881190030</td><td></td><td></td><td>2</td><td>2 days ago</td><td>●</td></tr>
</table>

I will make sure this test ID uses the validates="required" attribute to ensure the annotators don’t miss it.
I will make sure to provide tips why an answer is correct, incorrect, or unknown.
For the questions, I will include a free text comments field for annotators to use to explain and provide a reason for their answer.
I will review the images provided to the annotators to ensure I can classify them myself if I was presented the same questions I am asking them.
I will review with all the annotators the reason why they missed the question and interview an expert annotator to get help on improving the test question.</td>
</tr>
<tr>
<td><b>Contributor Satisfaction</b>

Say you’ve run a test launch and gotten back results from your annotators; the instructions and test questions are rated below 3.5, what areas of your Instruction document would you try to improve (Examples, Test Questions, etc.)</td>
<td>
<b>Contributor Satisfaction</b>
Number of participants: 20

3.2 / 5
Overall
<table>
<tr><td>3.3 / 5</td><td>2.9 / 5</td><td>2.8 / 5</td><td>3.7 / 5</td></tr>
<tr><td>Instructions Clear</td><td>Test Questions Fair</td><td>Ease Of Job</td><td>Pay</td></tr>
</table>

I think that reviewing the questions with subject matter experts may be biased itself because they are experts in the domain. I might review with non-SMEs as well.

The annotators should come from different geographies or backgrounds or cultures to normalize biases.</td>
</tr>
</table>

Limitations & Improvements

Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	The dataset contains 117 rows of images. To ensure non-biased results, the number of questions asked should evenly contain healthy, unhealthy, and inconclusive data sets. I used 1 test question for every 19 data points so that is about 6 questions needed. I would suggest using more data points for images. 118 may not be representative of all children patients. For example, I would use xrays from other hospital databases.
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	I would continually survey the human annotators to query their satisfaction with the test questions, ease of completing them, areas of improvement, and providing some incentive for helping improve the data labeling job like offering gift cards. I would submit a quick survey after the annotation jobs to get feedback. Responses as part of the test questions are different from post surveys which could garner more feedback about why annotators answered the way they did. I would use expert annotator as my control data when comparing with non-experts. I do this because I think expert annotators are biased.

Appendix A – CML code

```
<div class="row-fluid">
  <div class="span6">
    <img src="{{hosted_image}}"/>
  </div>
  <div class="span6">
    <cml:radios name="cloudiness_present" label="Do you see any areas of abnormal cloudiness/opacity in the lung?" 
validates="required">
      <cml:radio value="yes" label="Yes"/>
      <cml:radio value="no" label="No"/>
    </cml:radios>
    <cml:radios name="diapragm_present" label="Do you see a diaphragm shadow?" validates="required">
      <cml:radio value="yes" label="Yes"/>
      <cml:radio value="no" label="No"/>
    </cml:radios>
    <cml:radios name="heart_present" label="Do you see the heart?" validates="required">
      <cml:radio value="yes" label="Yes"/>
      <cml:radio value="no" label="No"/>
    </cml:radios>
    <cml:checkboxes name="cloudy_types" label="What area of the lung is opaque?" validates="required" only-
if="cloudiness_present:[yes]" exact="true">
      <cml:checkbox value="lhs" label="Big patch on left lung"/>
      <cml:checkbox value="rhs" label="Big patch on right lung"/>
      <cml:checkbox value="several" label="Several small cloudy areas on both sides"/>
    </cml:checkboxes>
    <cml:select name="confidence_level" label="How confident are you?" validates="required" only-
if="cloudiness_present:[yes]" exact="true">
      <cml:option label="1-not confident" value="1"/>
      <cml:option label="2-somewhat confident" value="2"/>
      <cml:option label="3-very confident" value="3"/>
    </cml:select>
    <cml:textarea name="confidence_detail" label="Confidence Level detail:" validates="required" only-
if="confidence_level:[2]" exact="true"/>
  </div>
</div>
```

Appendix B – Preview of Job

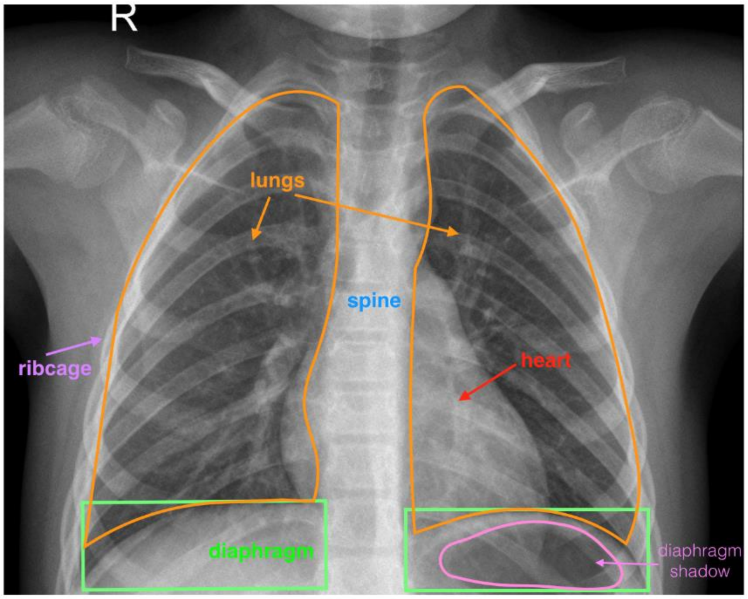
Overview

Help us determine if the images show a healthy lung or if the patient has pneumonia.

Steps

1. Examine the image and properly identify these annotated areas in the image.

- lungs
- spine
- heart
- ribcage
- diaphragm (below the lungs)



2. Check the appropriate box if the image shows clear lungs without any areas of abnormal cloudiness/opacity.

3. Check the appropriate box if the image shows the diaphragm shadow.

4. Check the appropriate box if the image shows the heart.

5. Indicate your confidence level in making your above choices from 1-3.

6. If you are somewhat confident (you picked confidence level 2 above), then type your reasons.

Rules Tips

Rules:

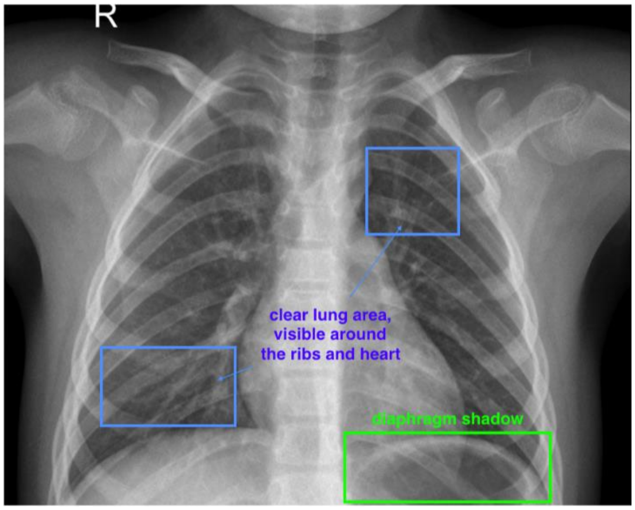
- Categorize cloudiness as either whole patches or small regions;
- If you are not confident in your answers, type the reason.

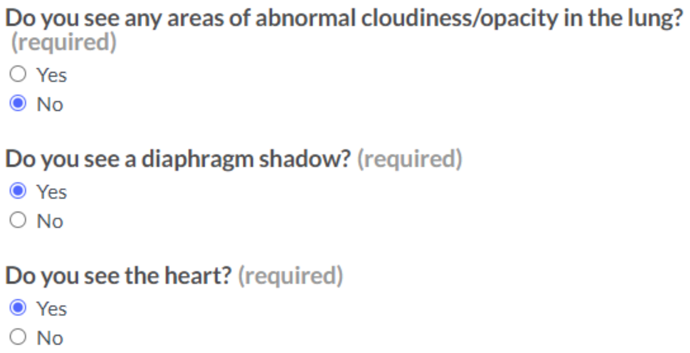
Tips:

- The R in every image indicates the right hand side. Please be careful when answering the checkbox indicating which side of the lung is cloudy.

Examples

Normal and healthy




Do you see any areas of abnormal cloudiness/opacity in the lung? (required)

○ Yes
● No

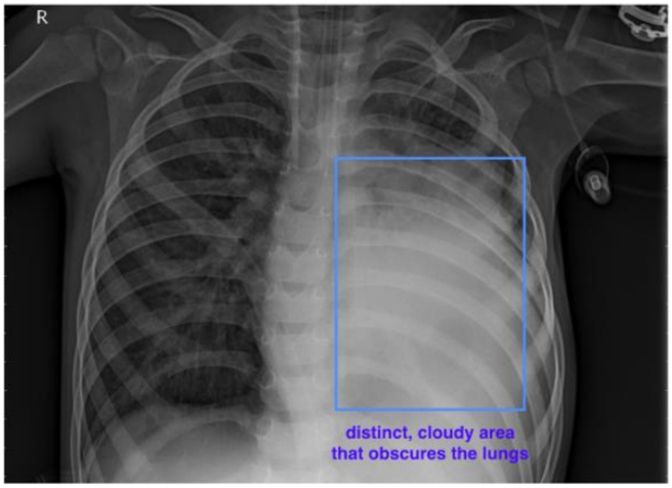
Do you see a diaphragm shadow? (required)

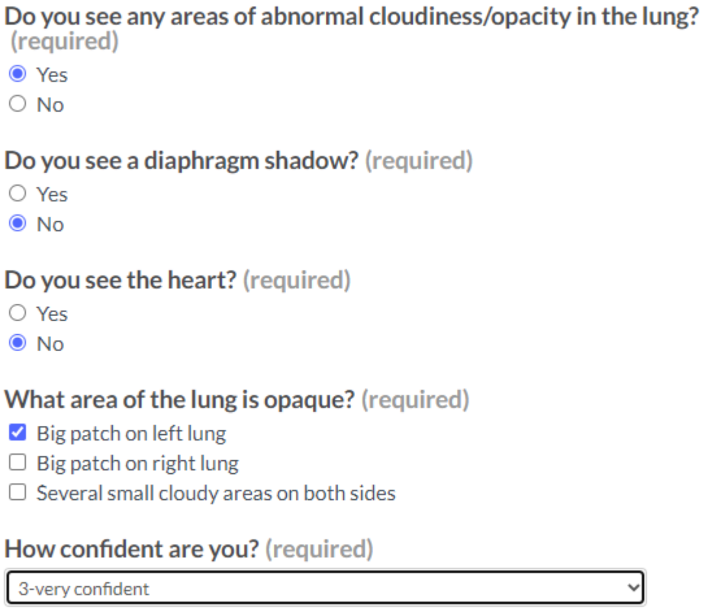
● Yes
○ No

Do you see the heart? (required)

● Yes
○ No

Abnormal with possible pneumonia




Do you see any areas of abnormal cloudiness/opacity in the lung? (required)

● Yes
○ No

Do you see a diaphragm shadow? (required)

○ Yes
● No

Do you see the heart? (required)

○ Yes
● No

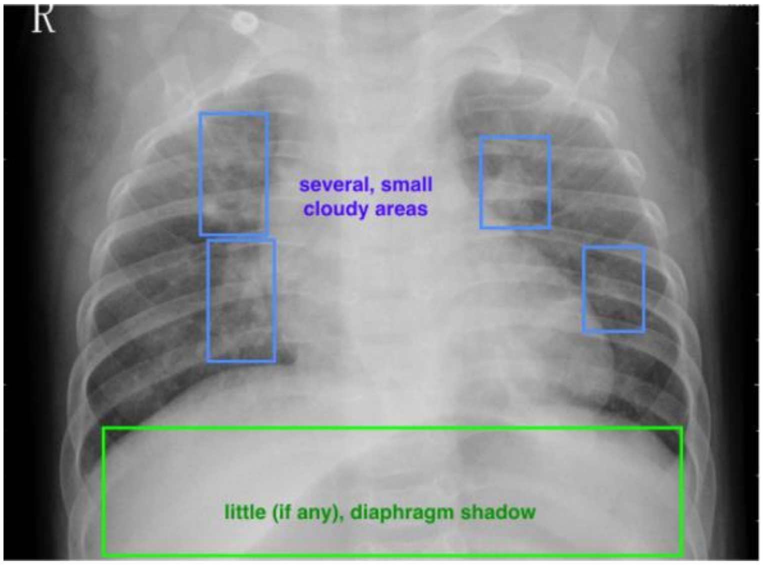
What area of the lung is opaque? (required)

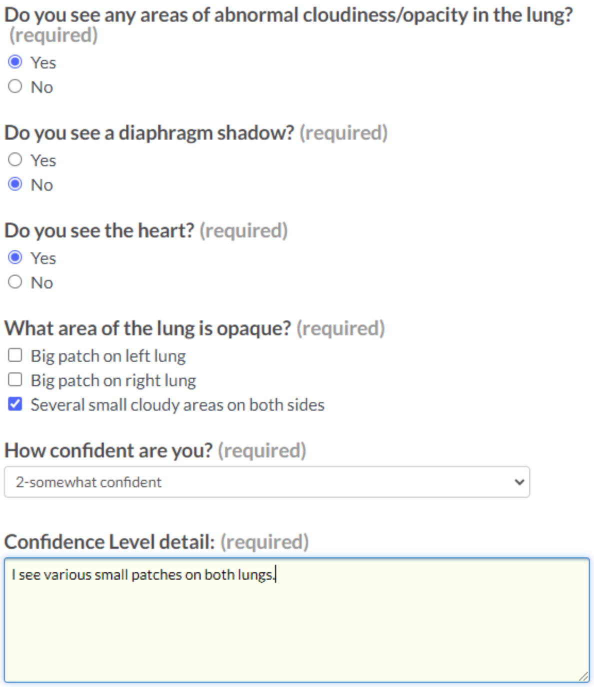
☑ Big patch on left lung
□ Big patch on right lung
□ Several small cloudy areas on both sides

How confident are you? (required)

3-very confident

Abnormal with possible pneumonia


R
several, small cloudy areas
little (if any), diaphragm shadow



Do you see any areas of abnormal cloudiness/opacity in the lung? (required)

● Yes
○ No

Do you see a diaphragm shadow? (required)

○ Yes
● No

Do you see the heart? (required)

● Yes
○ No

What area of the lung is opaque? (required)

☐ Big patch on left lung
☐ Big patch on right lung
☑ Several small cloudy areas on both sides

How confident are you? (required)

2-somewhat confident

Confidence Level detail: (required)

I see various small patches on both lungs.