

Project Proposal



Olivier Polack

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 problem is to distinguish x-rays of children lungs between those of children with pneumonia and those of healthy children. Machine learning will help screening out pneumonia x-ray images.
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?	<p>The labels I decided to add in my data are :</p> <p>Yes, this is a pneumonia x-ray image</p> <p>No, this is not pneumonia x-ray image</p> <p>Unsure</p> <p>Why?</p> <p>The goal is to help the annotator focus on all of the 'as objective as possible' observable criteria leading to the classification. So there must be a reminder of the visual criteria to observe and rate. Then these ratings lead to the label choice: Yes, No or Unsure. This important to let annotator leave a</p>

Test Questions & Quality Assurance

Number of Test Questions

Considering the size of this dataset, how many test questions did you develop to prepare for launching a data annotation job?

I prepared 3 checkboxes so annotators can judge if the x-ray image is displaying symptoms for sure, in which case they answer yes this is a pneumonia x-ray, or if not they answer No. But there is a third option as some images might not be as clearly displaying the symptoms they are looking for.

Improving a Test Question

Given the following test question which almost 100% of annotators missed, statistics, what steps might you take to improve or redesign this question?

ID	% CONTESTED	% MISSED	JUDGMENTS	LAST UPDATED	ENABLED
1881190030	<div><div></div></div>	<div><div></div></div>	2	2 days ago	<input checked="" type="checkbox"/>

I would look for ways to get direct feedback from few annotators about the fact they missed the question.

Contributor Satisfaction

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.)



Obviously "Ease of job" and "Test Questions Fair" have the lowest scores so those are the points to work on. "Ease of job" is only rated 2.8. This can mean there are too many answers to give, but most probably it means they find it hard to answer the questions. Perhaps what they have to spot in the pictures are for them too difficult to spot with confidence, in which case they need additional or better explanations beforehand with more illustrative examples. Showing a bit more examples could help, but surely finding out more consistent examples will help. "Test Questions Fair" is rated 2.9: Possibly the annotators would prefer to answer in a more detailed way about visible criteria, like for example "clear lungs", "cloudy areas", "diaphragm shadow not visible" and "diaphragm shadow visible" instead of making a final judgement.

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?

If the dataset contains much more samples of pneumonia x-rays than healthy x-rays, then the bias will be towards x-ray pneumonia.

The data might be improved if we make sure the x-images used are from the same pool of images on which the tool is aimed at working on. If the images used for the dataset are coming from children already in hospitals and with lungs issue then the bias might be towards detecting false positive if used on random x-rays from children lungs.

The dataset could be improved if the pool is matches the usual percentage of pneumonia versus healthy cases found usually.

Designing for Longevity

How might you improve your data labeling job, test questions, or product in the long-term?

Potential stakeholders (engineers, medical professionals, etc.) discussions that could help

Ask medical professionals if more objective criteria to added.

Ask engineers if relative transparencies of areas can be monitored and might become meaningful objective criteria

Are there not any other specific visible symptoms or situations that should be annotated, symptoms of other diseases or simply abnormalities that are a potential health issue?