Project Proposal

Antonina Savka

Data Labeling Approach

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML in solving this task?	One of the first crucial steps in pneumonia diagnoses is an analysis of the patience's lungs X-Ray scan. A professional radiologist or pulmonologist can determine the symptoms by looking at the scan within the second. The aim of this project to help doctors identify the symptoms of pneumonia in children on X-Ray scans of the lungs. As result, doctors have more time for other activities that require more human intelligence. To active the aim, we can use an AI system that categorises the images of the scans with appropriate labels. And with a well-trained AI system, we can target the better accuracy of diagnosis and prevent them from the human-errors.

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?	The labelling schema contains the 3 labels: • Yes - for the images with pneumonia • No - for the images without pneumonia • Unknown - to address uncertainty At this stage of the project, we want to achieve from AI a basic understanding of pneumonia. That is why we keep the knowledge simple yet just by labelling it as Yes/No (pneumonia or not pneumonia). The Unknow label is not a desirable label but it helps us to detect the problematic set of data: if the user cannot easily classify the image, the AI will likely make a mistake as well in the early stages of the project. That is why it is important to detect such data to improve instructions on the next iteration of the labelling and learning and address the ambiguity. **Additional information** To get more knowledge about the provided data set (whether it is suitable and accurate), additional information is requested from the annotators. When the annotator classifies the image as Yes (with pneumonia), they are asked to provide additional information such as: • The severity of pneumonia: ○ On early stage ○ Significant ○ Severe • The symptoms on the image that allows the user to classify it as pneumonia: ○ Areas of cloudiness/opacity in the lungs ○ Little or any diaphragm shadow The severity level will help us with 2 aspects: 1. We can use this information on the following iteration to extend the range of the labels and, as result, to help AI make a more complex decision: how severe the pneumonia is. 2. For the images with the mistaken label or ambiguity, we can detect why the annotator decided that the image shows pneumonia and, as result, improve the instructions. The choice of symptoms will help with these aspects: 1. We can better understand why the annotator chooses to set the label Yes. If the label is incorrect, this information will show us what area of instructions has to be improved. Prevent selecting the Yes label by mistake: the user is required then to provide the additional information, so we know that the Yes option was selected consciously

Test Questions & Quality Assurance

<table>
<tr>
<td>

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?

</td>
<td>

The test plan contains 15% of the data set (16 test files and 101 not labeled file).
The key focus is on the Yes/No labels tests distribution. The Unknow label is not a desirable outcome, despite its value in determining ambiguity and data set accuracy. As the annotation job is not free, we tried to eliminate the invalid or ambiguous data from the data set as much as we could. That is why the Unknow label test coverage is less important.
The distribution of the tests:

- About 50% on the “No” pneumonia label
- About 50% on the “Yes” pneumonia label:
  - About 17% on moderate severity (“On early stage”)
  - About 17% on high severity (“Significant”)
  - About 17% on critical severity (“Severe”)
- “Unknown” may not be accounted for in testing as this label is for analysing data set quality

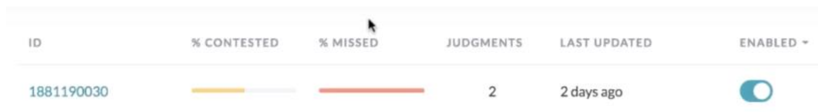
The distribution of the symptoms is not valuable for testing as this is additional information for analysing the quality of the instruction.

</td>
</tr>
<tr>
<td>

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?

</td>
<td>



1. Review test case. Did we do mistake when we provided the control answer? Or is it possible to apply the rules describe in the instruction to this test?
2. Revisit the rules provided in the instruction. It is likely that we need to improve the rules description: make it more detailed. It is highly recommended to consult with the specialist (e.g. radiologist or pulmonologist) to address the uncertainty of the rules.
3. Revisit the examples provided in the instruction. The examples set may not the case we are asking about in the test.

</td>
</tr>
</table>

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.)

These 3 indicators correlate with each other. It give the idea that the problem may lay in these areas:
1. Not clear rules of detecting the label. In there is are difficulties to explain the rules, I would suggest involving a specialist (e.g. radiologist or pulmonologist) to help to clarify the rules of identifying the pneumonia on the scan.
2. Revisit the dataset in terms of accuracy: if the dataset contains the scans from the different perspectives (e.g. from the side instead of the front) the rules simply are not applicable to some of the images.
3. Revisit the examples and test set: it is possible that we missed some edge cases in the examples and/or in the test set. In this case the annotator learned based on one information but was asked to judge the different information.

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?	1. Most of the images contain the pneumonia symptoms in the left lung. We may need to extend the dataset to cover right-side pneumonia as well as the pneumonia of the top part of the lungs. 2. All scans are well made front X-Rays. The emergency scans may contain arms shadows, artefacts (e.g. tubes) etc. These additional “objects” on the scan can lead to mislabelling. The test cases are created based on the unprofessional judgment of the team members. This may lead to teaching an AI model based on invalid data. Involving the professional (e.g. radiologist or pulmonologist) as a consultant is **highly recommended.**
Designing for Longevity How might you improve your data labeling job, test questions, or product in the long-term?	Using the feedback from the annotators we should: 1. Revisit the dataset in terms of how accurate the data is. For example, are all images taken from the front or some of them are side scans? In the last case, it will be difficult to detect the symptoms of pneumonia. 2. Revisit the rules provided in the instruction. It is likely that we need to improve the rules description: make it more detailed. It is highly recommended to consult with the specialist (e.g. radiologist or pulmonologist) to address the uncertainty of the rules. 3. Revisit the examples provided in the instruction. The examples set may not cover all edge cases. Revisit the labels. We may want to include the severity in the main label set.