Weak Supervision

Vincent Chen and Nish Khandwala

Outline

- Motivation
    - We want more labels!
    - We want to “program” our data! *#Software2.0*
- Weak Supervision Formulation
- Landscape of Noisy Labeling Schemes
- Snorkel Paradigm
- Demos
    - Writing labeling functions (LFs) over images
    - Cross modal

Problem 1: We need massive sets of training data!

Massive sets of **hand-labeled data** → **Modern supervised learning** (e.g. our beloved ConvNets!)

- High cost + inflexibility of hand-labeled sets!
  - Medical Imaging: How much would it cost for a cardiologist to label thousands of MRIs?

Problem 1: We need <u>massive</u> sets of training data!

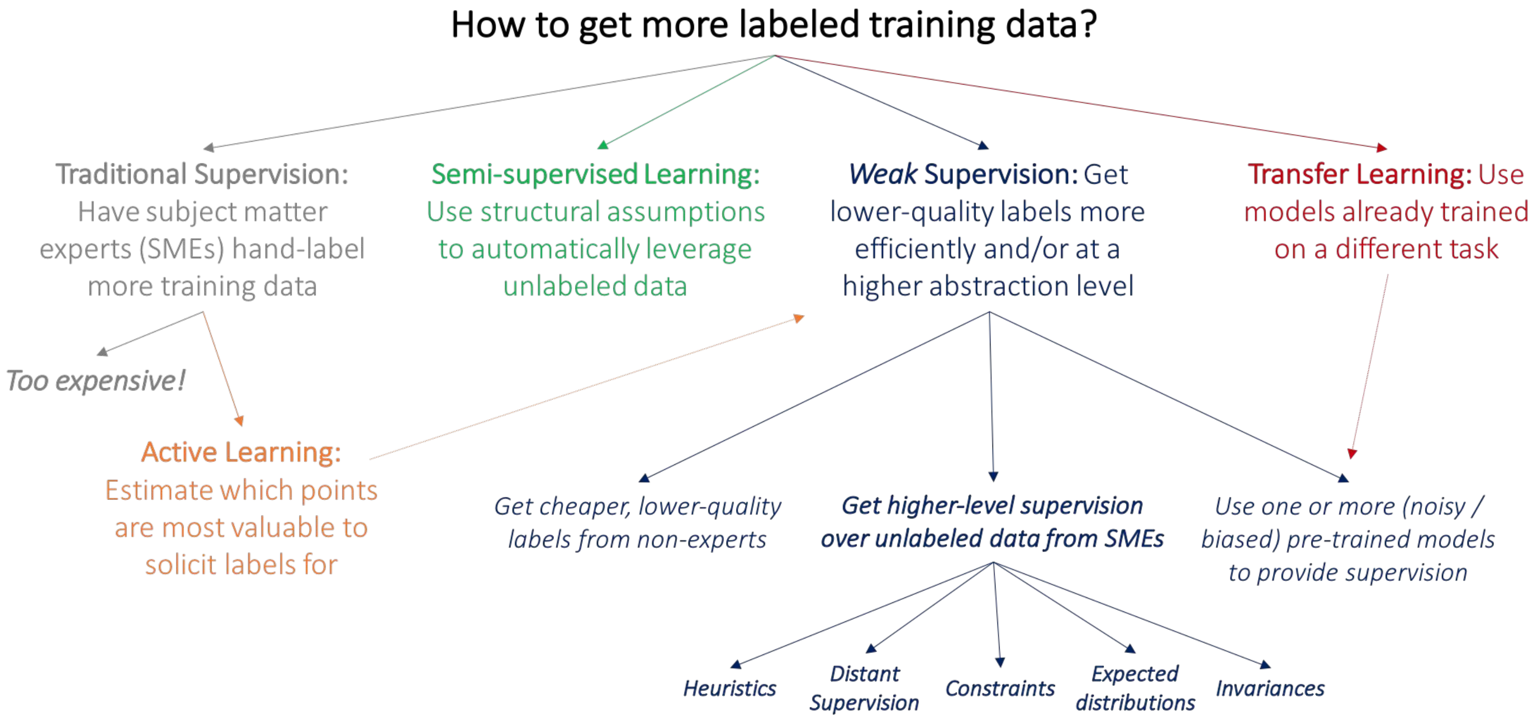


Image: https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/

Problem 2: We want to *program* our data with domain expertise!

- Software 2.0: biggest challenge is **shaping your training data**!
- Weak supervision as an approach to **inject domain expertise**

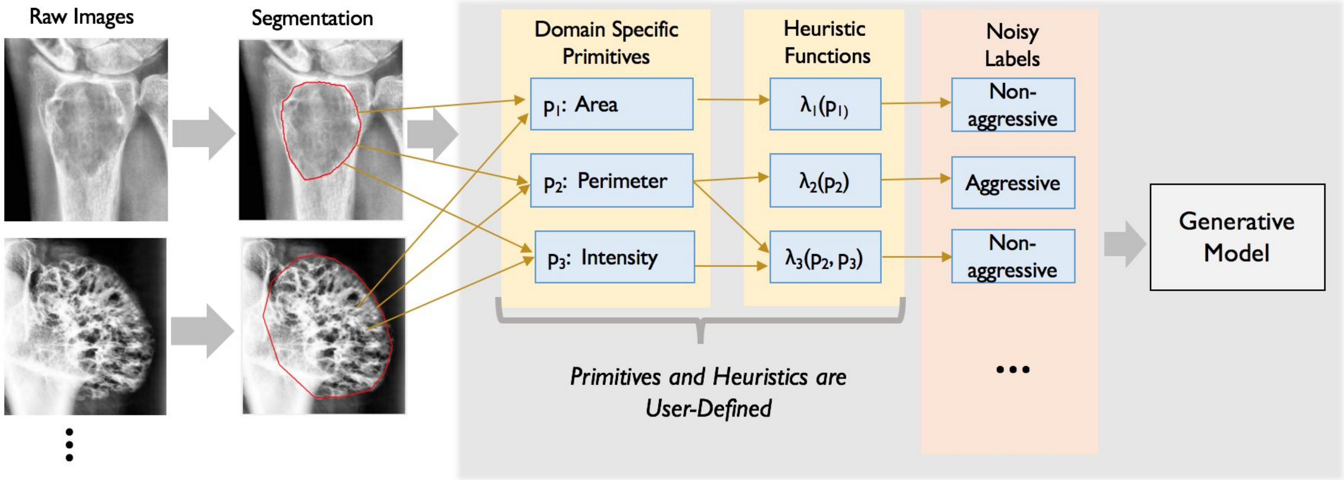


Figure: Varma et. al 2017 https://arxiv.org/abs/1709.02477

Problem 2: We want to *program* our data with domain expertise!

Programming by curating noisy signals!

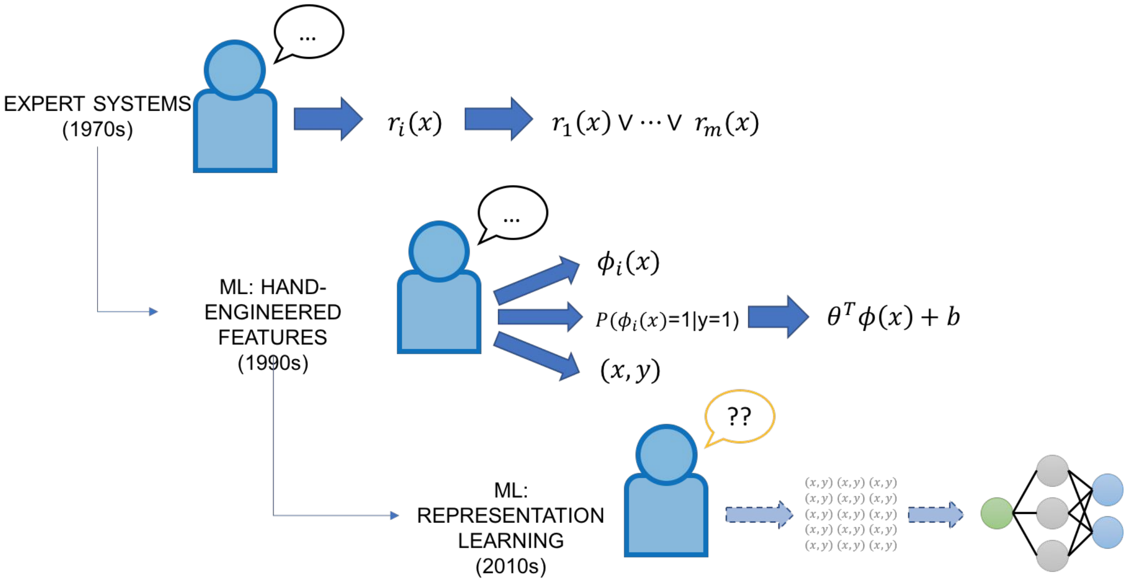


Image: https://hazyresearch.github.io/snorkel/blog/snorkel_programming_training_data.html

Weak Supervision Formulation

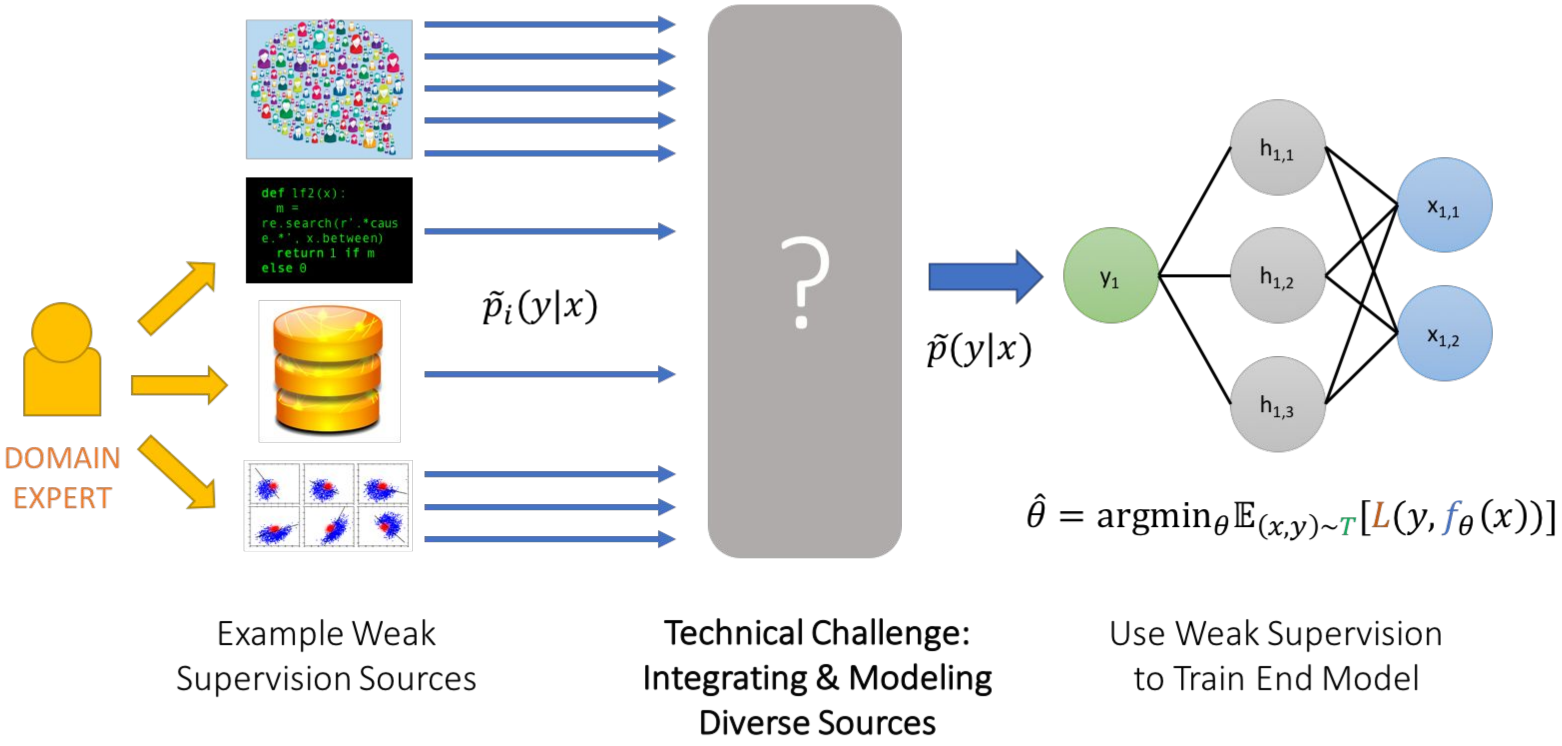
However, instead of having ground-truth labeled training set, we have:

- Unlabeled data, $X_u = x_1, \ldots, x_N$
- One or more weak supervision sources of the form $p'_i(y \mid x)$, $i = 1 : M$, provided by a human domain expert such that each one has:
    - A coverage set, C_i, the set of points x over which source is defined
    - An accuracy, defined as the expected probability of the true label, y^* over its coverage set, which we assume is < 1.0

- Learn a <u>**generative model**</u> over <u>coverage</u> and <u>accuracy</u>

Source: A. Ratner et. al https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/

Weak Supervision Formulation

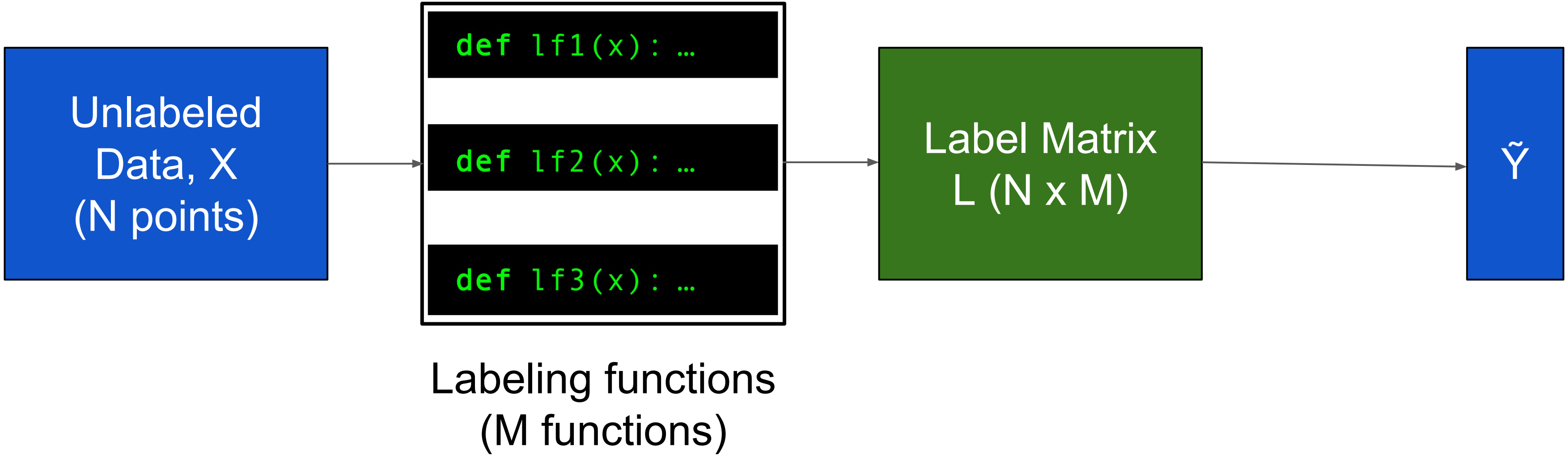


Source: A. Ratner et. al https://dawn.cs.stanford.edu/2017/07/16/weak-supervision/

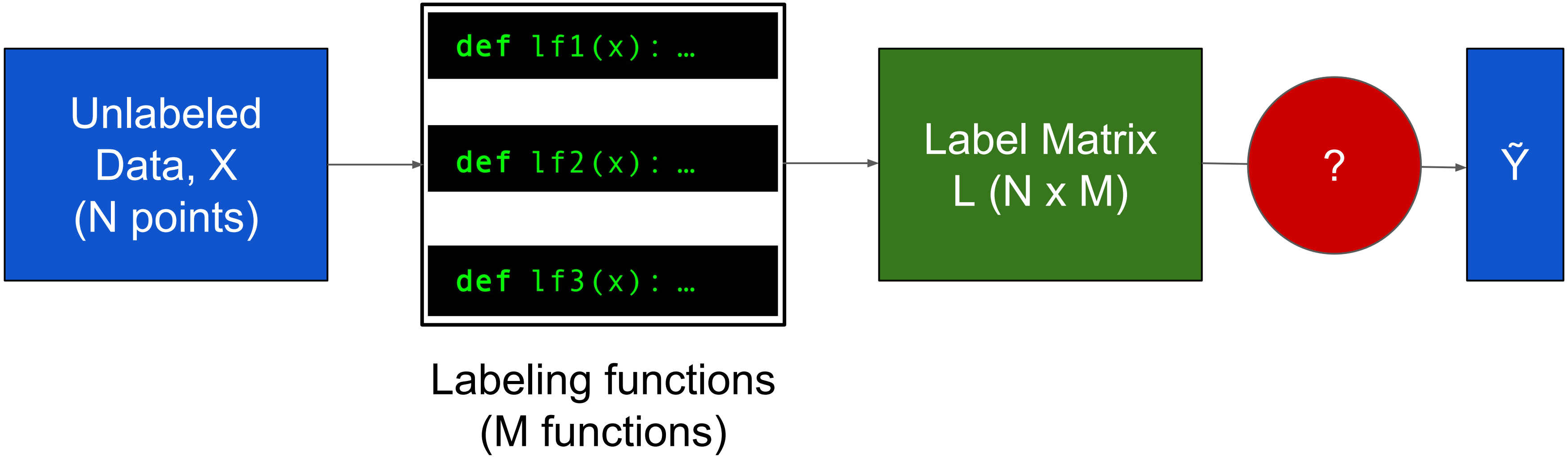
Data Programming

- Recent method proposed by Alex Ratner from Prof. Chris Re’s group
- Composed of three broad steps:
    - Rather than hand-labeling training data, **write multiple labeling functions (LFs)** on X using patterns and knowledge bases
    - **Obtain noisy probabilistic labels, $\tilde{Y}$** --- how?
    - **Train an end model on X, $\tilde{Y}$** using your favorite machine learning model

Data Programming



Data Programming



Data Programming

How do we obtain probabilistic labels, $\tilde{\mathbf{Y}}$**,** from the label matrix, **L**?

Approach 1 - Majority Vote

Take the majority vote of the labelling functions (LFs).

Let’s say **L** = [[0, 1, 0, 1, 0]; [1, 1, 1, 1, 0]].

$\tilde{\mathbf{Y}}$ = [0, 1]

But this approach makes several strong assumptions about the LFs...

Data Programming

How do we obtain probabilistic labels, $\tilde{\mathbf{Y}}$**,** from the label matrix, **L**?

Approach 2

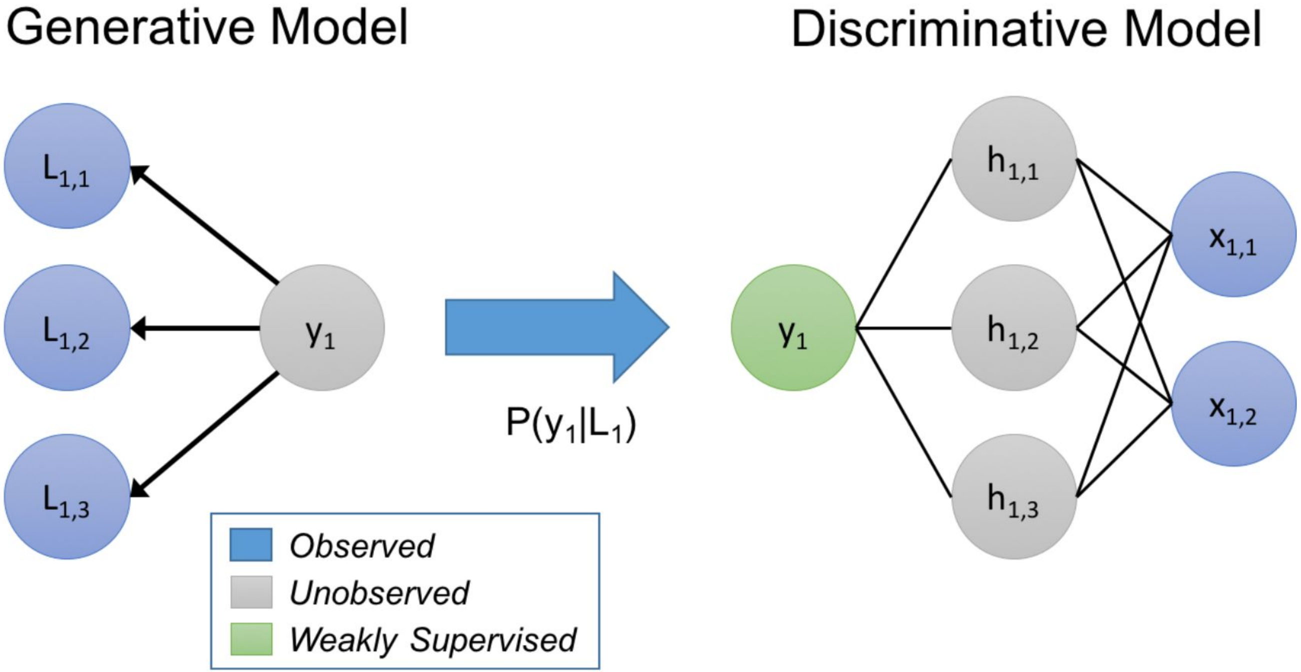
We train a generative model over **P(L, Y)** where **Y** are the **(unknown)** true labels. Recall from CS109 that **P(L, Y) = P(L | Y)P(Y)** → we don't need to know the true labels, **Y**!

$\tilde{\mathbf{Y}}$ can be obtained by taking a weighted sum of LFs' outputs, where the weights for the LFs are obtained from the generative model training step.

Intuition?

Data Programming

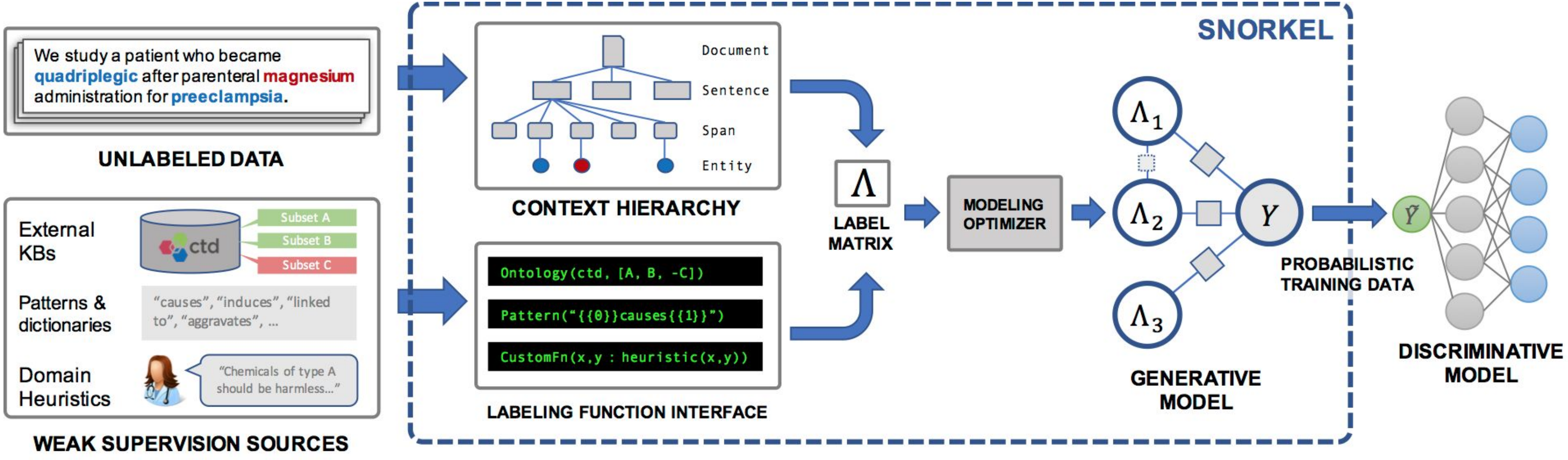
Putting it all together...



Source: A. Ratner et. al https://hazyresearch.github.io/snorkel/blog/weak_supervision.html

Data Programming

Putting it all together...



Source: A. Ratner et. al, Snorkel: Rapid Training Data Creation with Weak Supervision

Data Programming

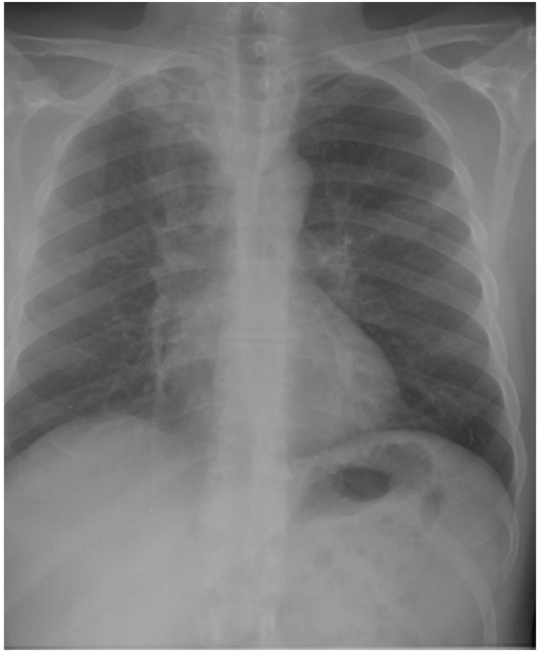
Framework available on GitHub: https://github.com/HazyResearch/snorkel

Demo: Writing LFs over Images

Tutorial: https://github.com/vincentschen/snorkel/blob/master/tutorials/images/Intro_Tutorial.ipynb

Let's write LFs for this image?

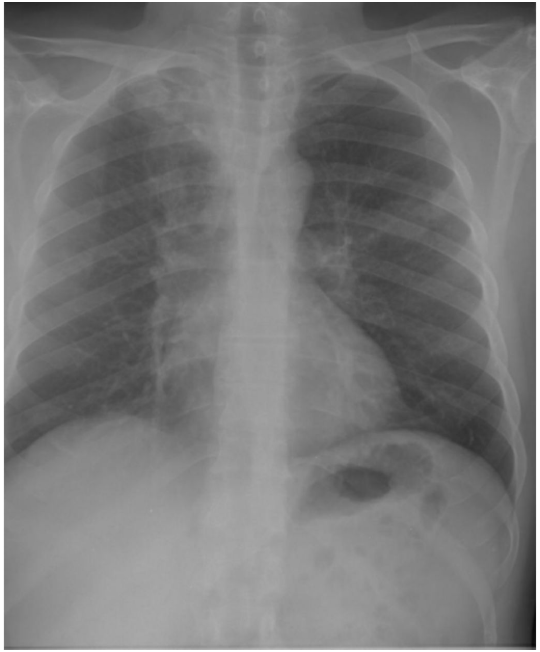
Task: Build a chest x-ray normal-abnormal classifier



Source: Open-I NLM NIH Dataset

How about now?

Task: Build a chest x-ray classifier



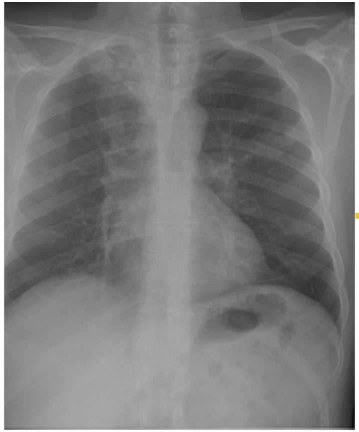
Indication: Chest pain. Findings: Mediastinal contours are within normal limits. Heart size is within normal limits. No focal consolidation, pneumothorax or pleural effusion. Impression: No acute cardiopulmonary abnormality.

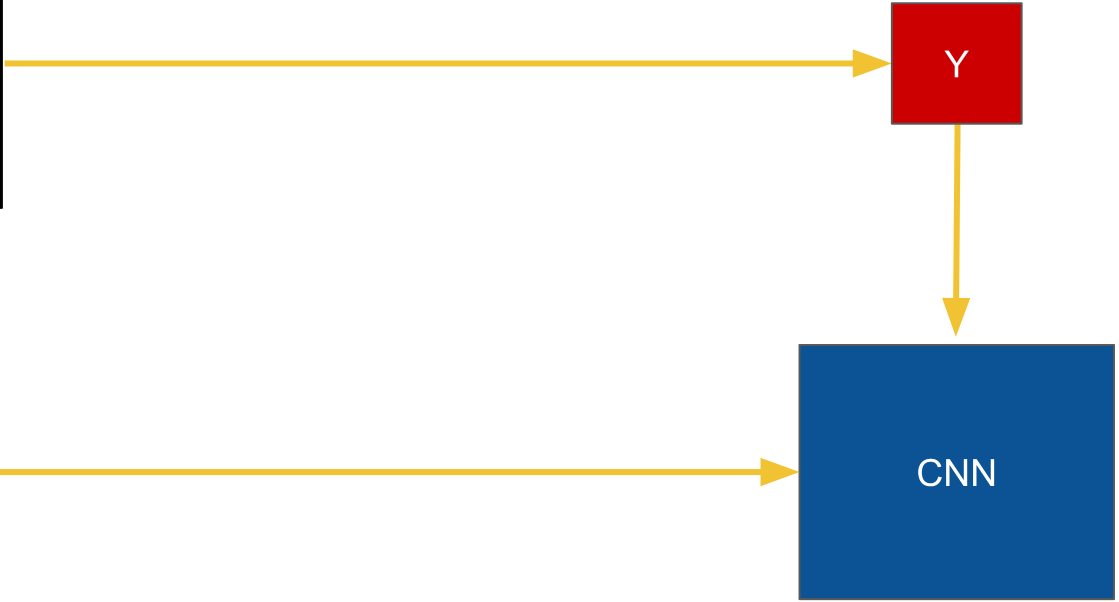
Can you use the accompanying medical report (text modality) to label the x-ray (image modality)?

This setting is what we call “cross-modal”!

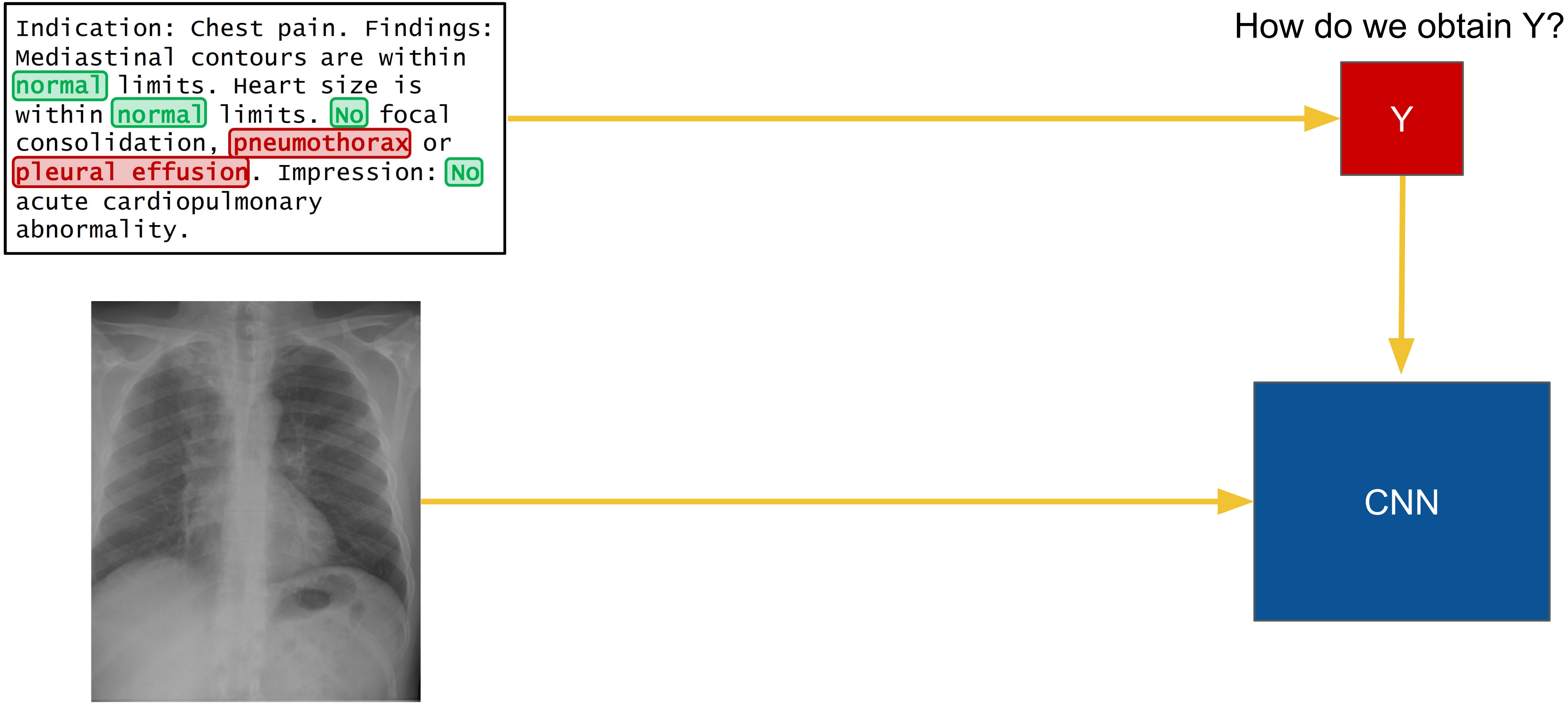
Cross-Modal Weak Supervision

Indication: Chest pain. Findings: Mediastinal contours are within normal limits. Heart size is within normal limits. No focal consolidation, pneumothorax or pleural effusion. Impression: No acute cardiopulmonary abnormality.





Cross-Modal Weak Supervision



Cross-Modal Weak Supervision

Indication: Chest pain. Findings: Mediastinal contours are within normal limits. Heart size is within normal limits. No focal consolidation, pneumothorax or pleural effusion. Impression: No acute cardiopulmonary abnormality.

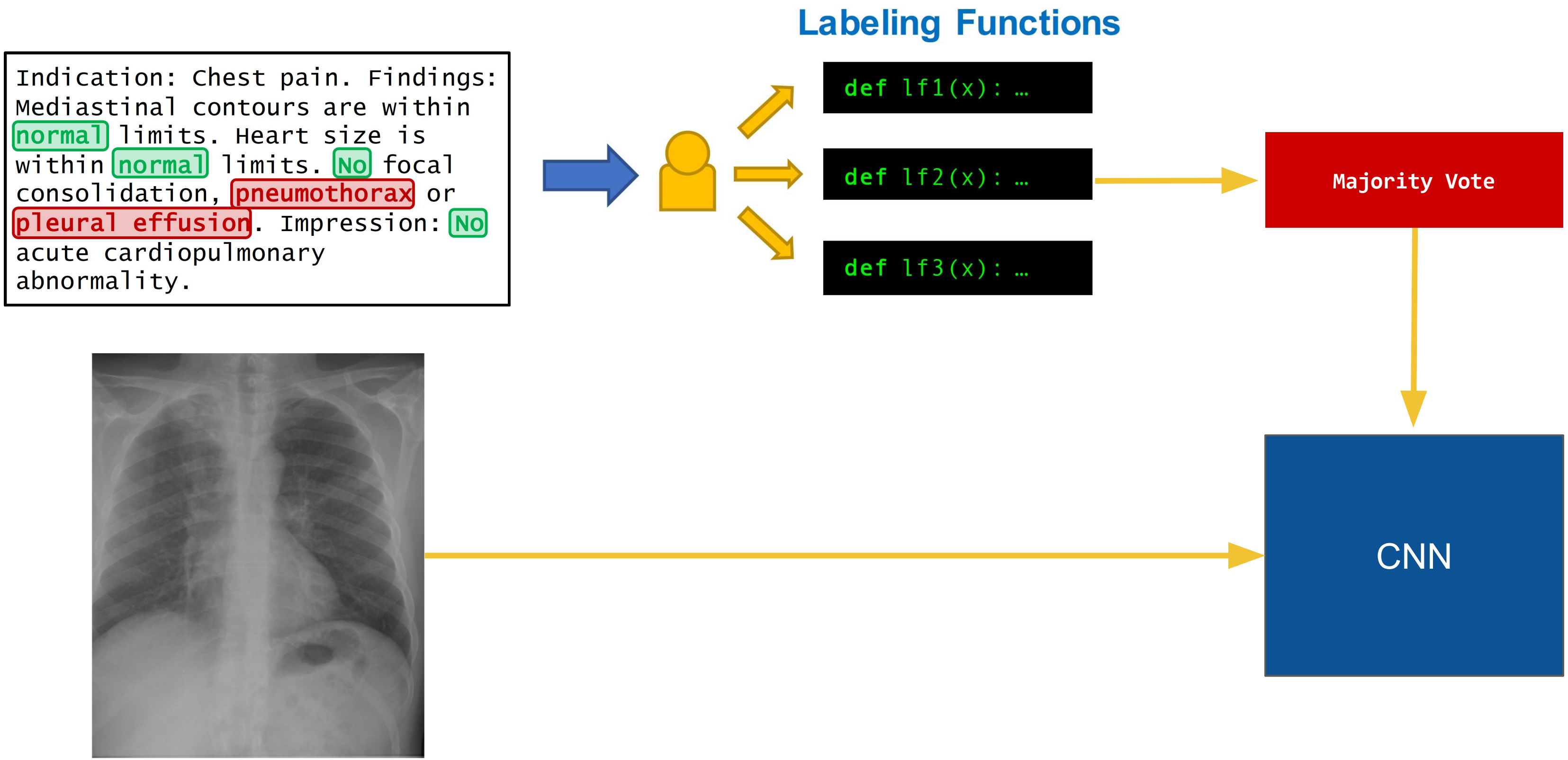
Normal Report

```
def LF_pneumothorax(c):
  if re.search(r'pneumo.*', c.report.text):
      return "ABNORMAL"

def LF_pleural_effusion(c):
  if "pleural effusion" in c.report.text:
      return "ABNORMAL"

def LF_normal_report(c, thresh=2):
    if len(NORMAL_TERMS.intersection(c.
    report.words)) > thresh:
      return "NORMAL"
```

LFs

Source: Khandwala et. al 2017, Cross Modal Data Programming for Medical Images

Cross-Modal Weak Supervision - Approach 1



Cross-Modal Weak Supervision

```
Indication: Chest pain. Findings:
Mediastinal contours are within
normal limits. Heart size is
within normal limits. No focal
consolidation, pneumothorax or
pleural effusion. Impression: No
acute cardiopulmonary
abnormality.
```

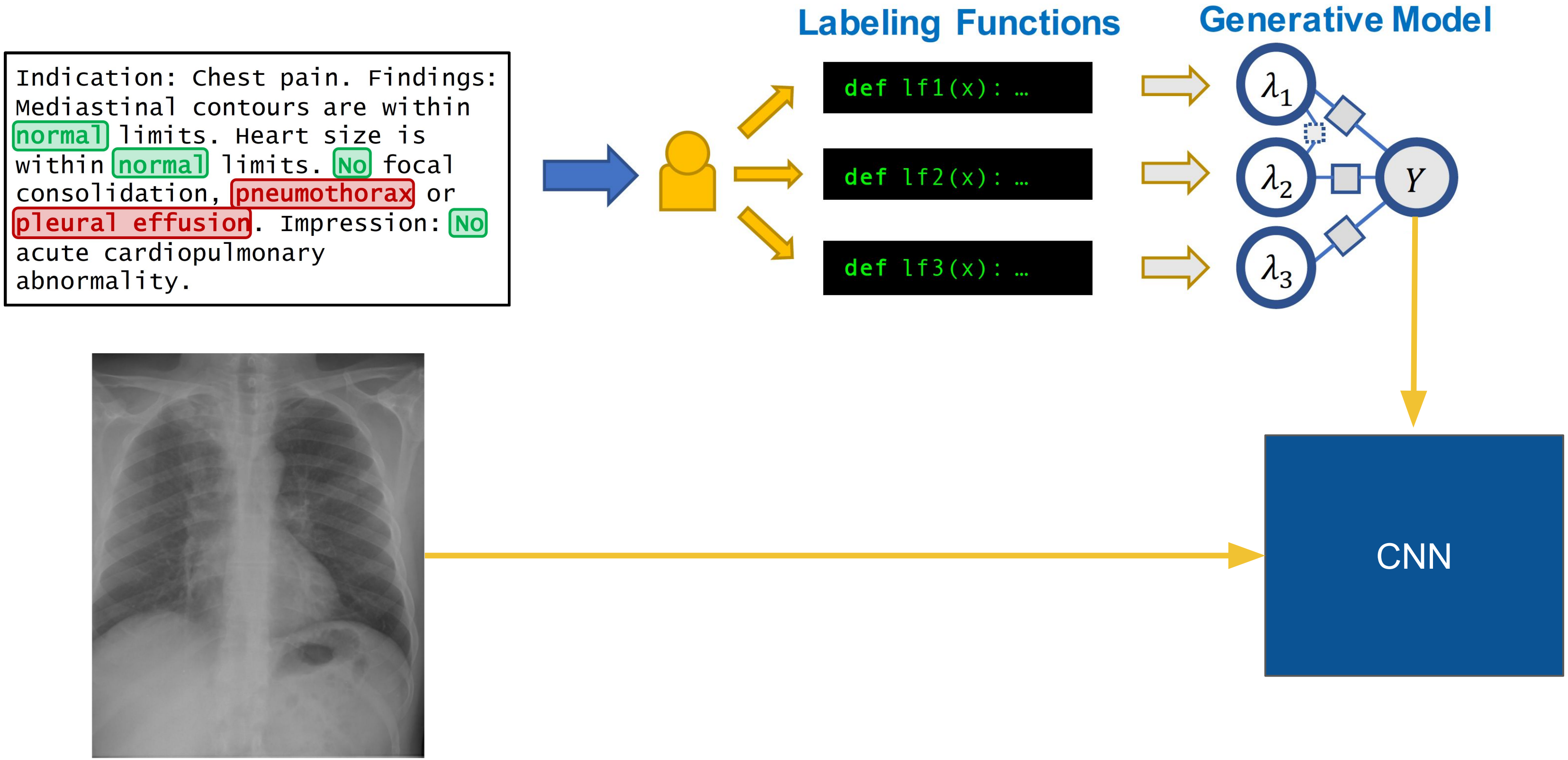
Normal Report

```
def LF_pneumothorax(c):
  if re.search(r'pneumo.*', c.report.text):
      return "ABNORMAL"

def LF_pleural_effusion(c):
  if "pleural effusion" in c.report.text:
      return "ABNORMAL"

def LF_normal_report(c, thresh=2):
    if len(NORMAL_TERMS.intersection(c.
    report.words)) > thresh:
      return "NORMAL"
```

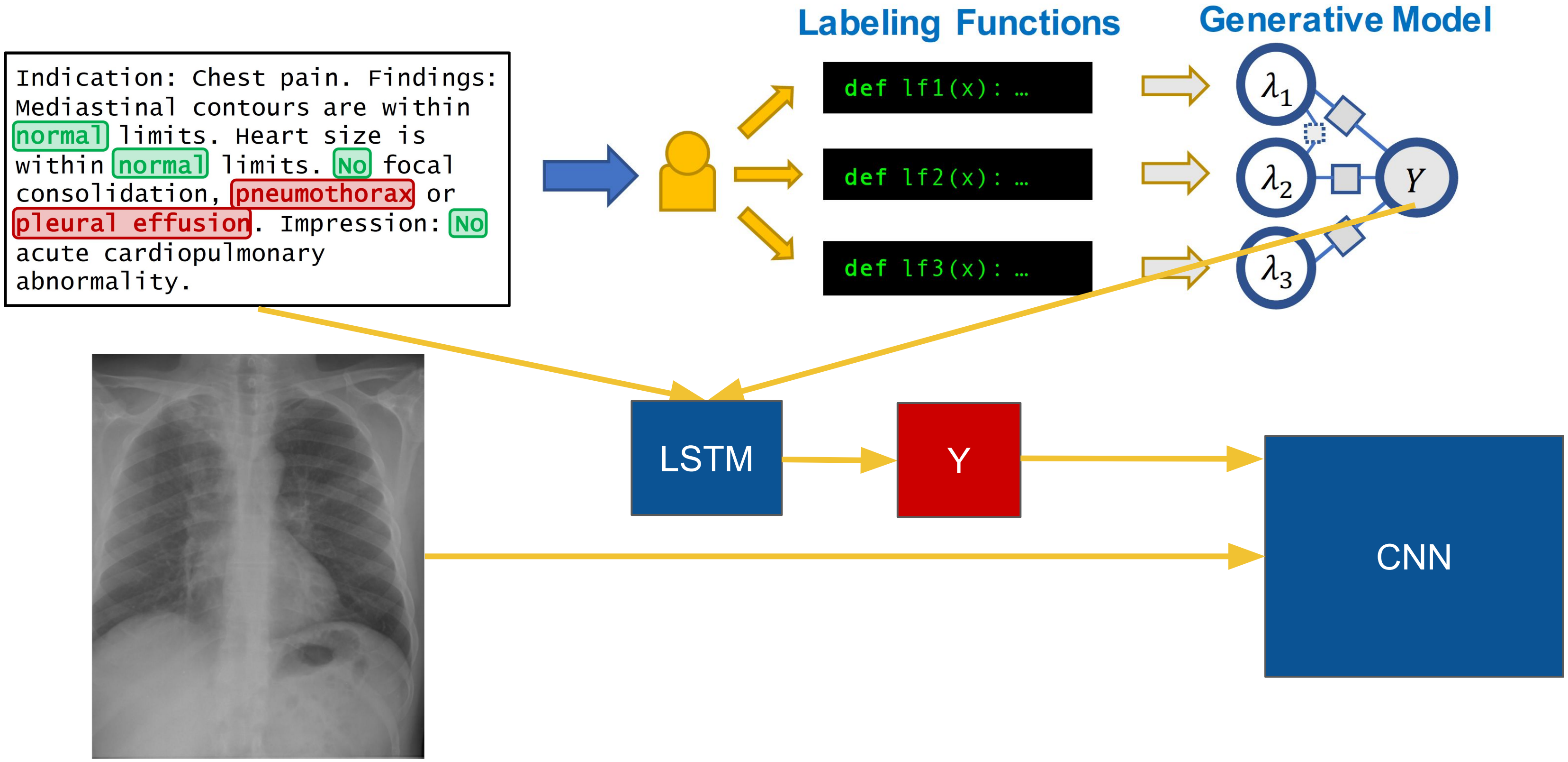
LFs

The first two LFs check for abnormal disease terms (in red), and the third LF checks for normal terms (in green). Here, Majority Vote (MV) outputs an incorrect abnormal label, but the Generative Model (GM) learns to re-weight the LFs such that the report is correctly labeled as normal.

Cross-Modal Weak Supervision - Approach 2



Cross-Modal Weak Supervision - Approach 3



How good are the labels?

Approach 1 (MV)	Approach 2 (GM)	Approach 3 (DM)
0.75	0.90	0.93

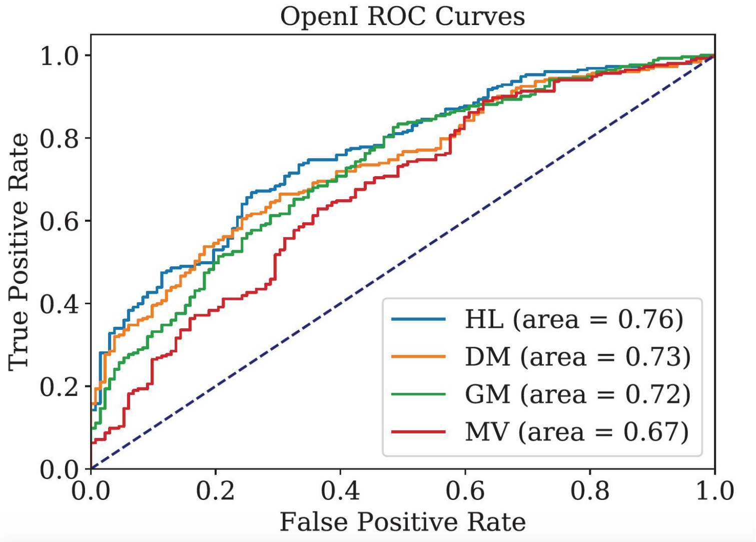
Test set AUC ROC scores (Open-I Chest X-ray Dataset)

Source: Khandwala et. al 2017, Cross Modal Data Programming for Medical Images

How good is the image classifier?

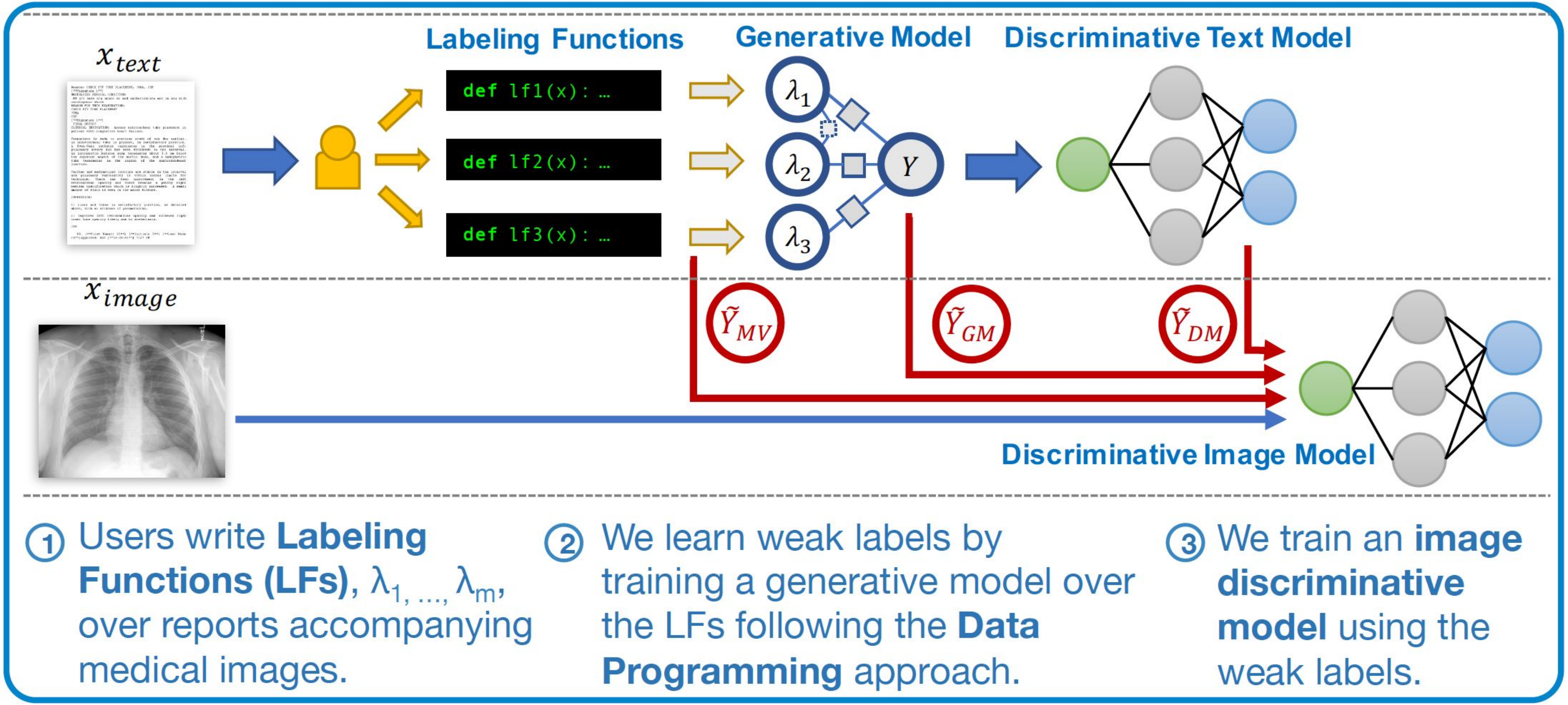
Approach 1 (MV)	Approach 2 (GM)	Approach 3 (DM)	Fully Supervised (HL)
0.67	0.72	0.73	0.76

Test set AUC ROC scores (Open-I Chest X-ray Dataset)



Source: Khandwala et. al 2017, Cross Modal Data Programming for Medical Images

Cross Modal Weak Supervision - Summary



Source: Khandwala et. al 2017, Cross Modal Data Programming for Medical Images