Automatically Annotate Medical Images for downstream analysis

Capstone Project (Spring 2022)

Li Yuan

Master of Science in Data Science

Data Science Institute, Vanderbilt University

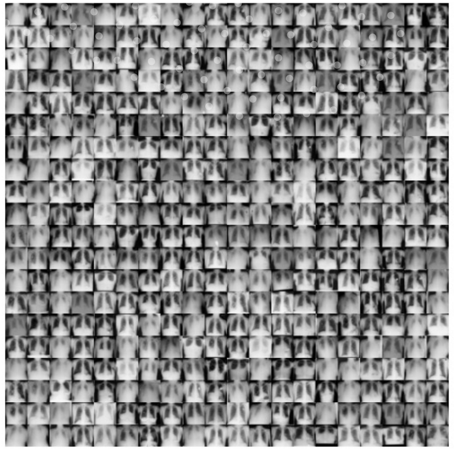
The Goal/Motivation

More and more unlabeled medical images are needed to be annotated for the downstream research analysis. However, manually labeling millions of medical images is impractical. Thus, I want to build an automatic annotator to speed up this process and provide more labeled images for big data analysis.

Data Description

MedMNIST v2: A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification

Facts of **ChestMNIST**



Data Modality: Chest X-ray

Task: Multi-Label (14) Binary-Class (

Number of Samples: 112,120 (78,46

Source Data:

Xiaosong Wang, Yifan Peng, et al., "Chest
weakly-supervised classification and localizat

License: *CC BY 4.0*

Facts of **BreastMNIST**

Data Modality: Breast Ultrasound

Task: Binary-Class (2)

Number of Samples: 780 (546 / 78 /

Source Data:

Walid Al-Dhabyani, Mohammed Gomaa, et al
104863, 2020.

License: *CC BY 4.0*

Design Modeling

Three Stages:

1. Directly Apply zero-shot clip model to classify the binary task for recognizing chest or breast images
2. Train a Convolutional Neural Network to fine tune predictions from the first step
3. Compare the trained CNN model results with results from CLIP on held-out test set

Model Architecture Pipeline

1. Select half chest images and half breast images
2. Stratified Split it into training, and test set

Model Architecture Pipeline

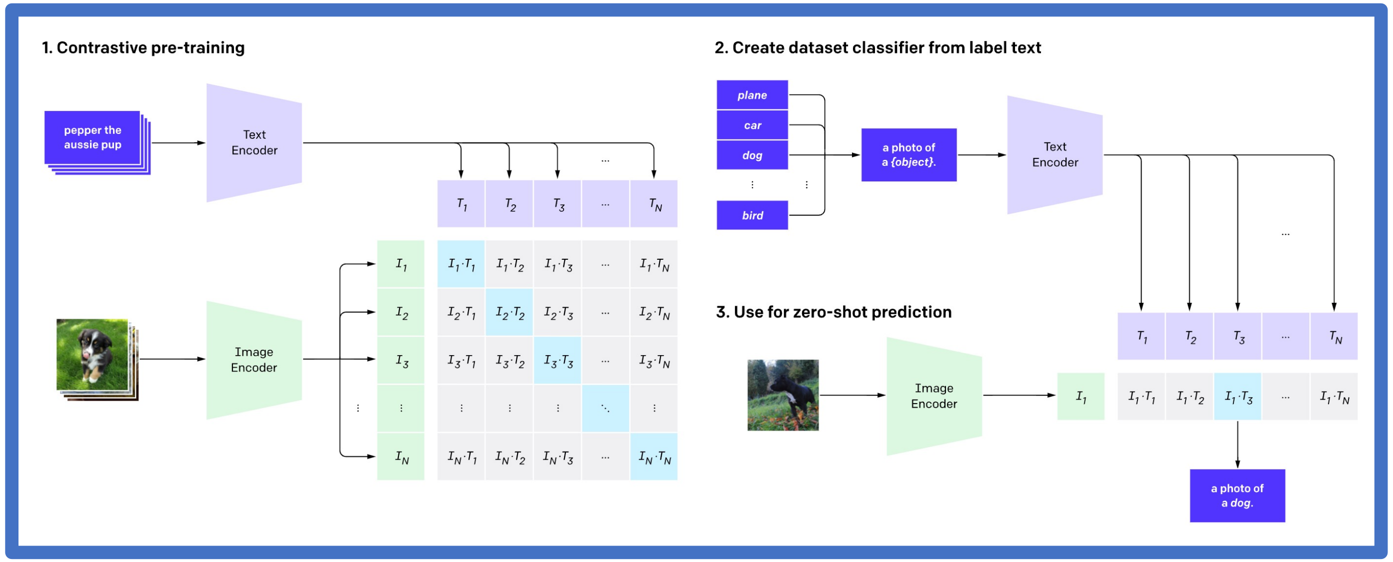
1. Select half chest images and half breast images
2. Stratified Split it into training, and test set

3. Feed training into CLIP

Model Architecture Pipeline

1. Select half chest images and half breast images
2. Stratified Split it into training, and test set

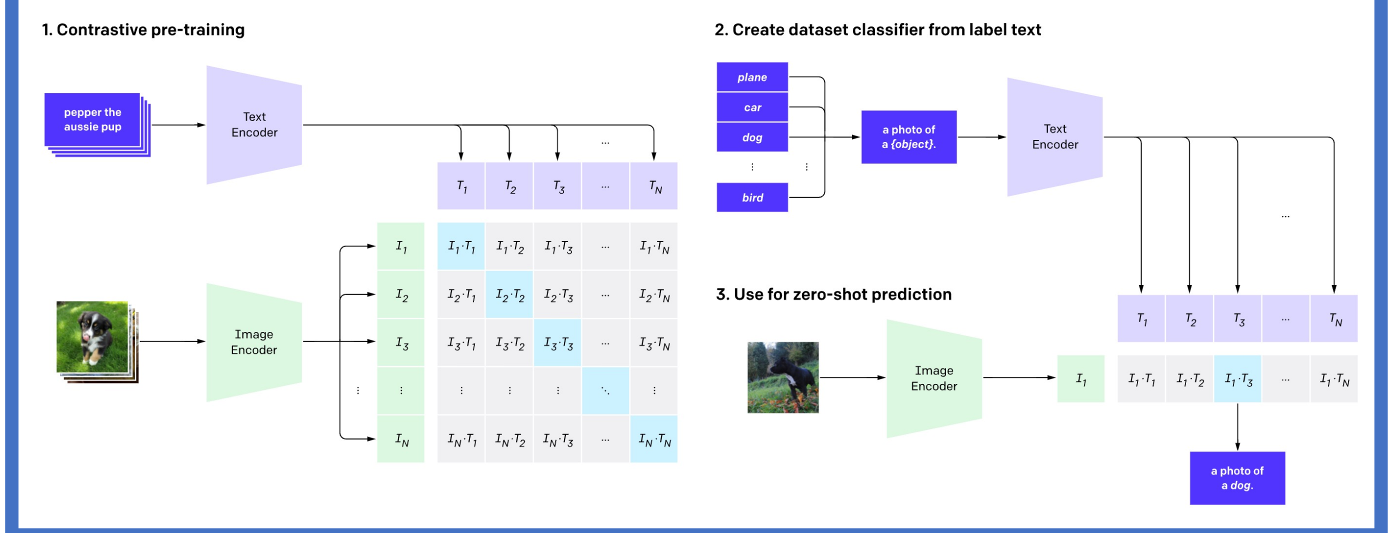
3. Feed training into CLIP



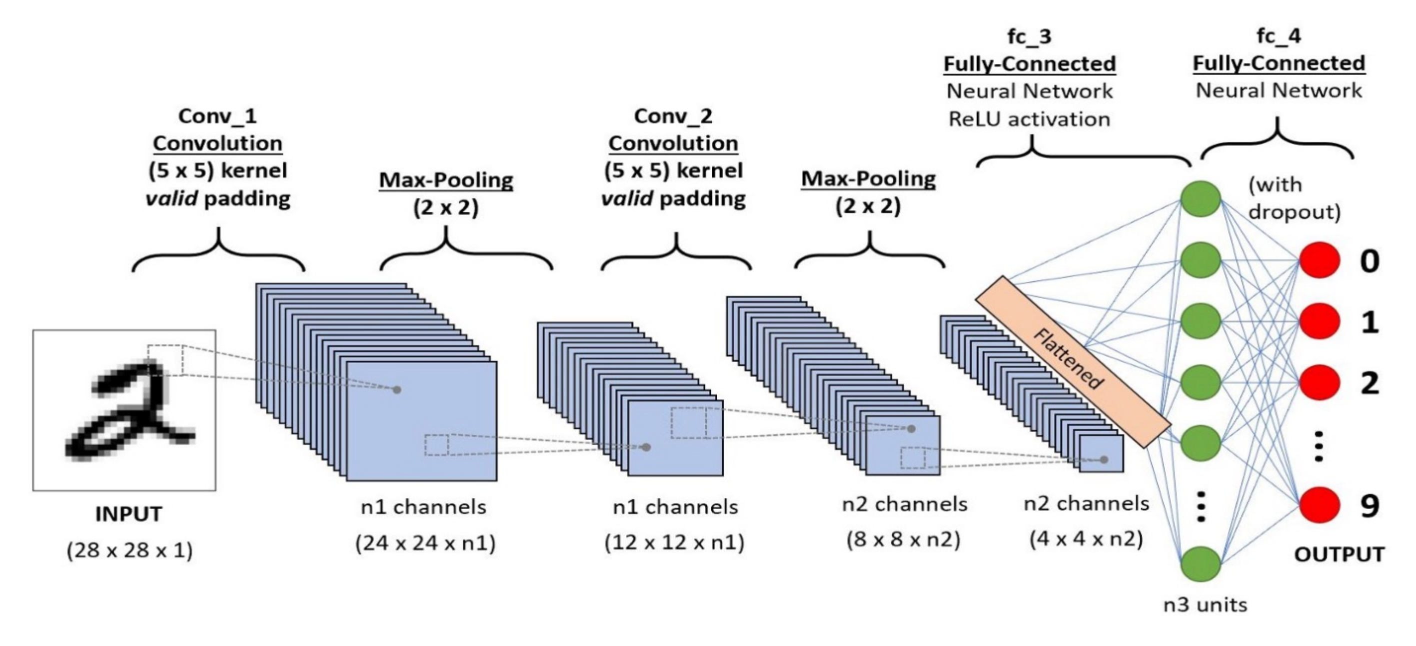
Model Architecture Pipeline

1. Select half chest images and half breast images
2. Stratified Split it into training, and test set

3. Feed training into CLIP



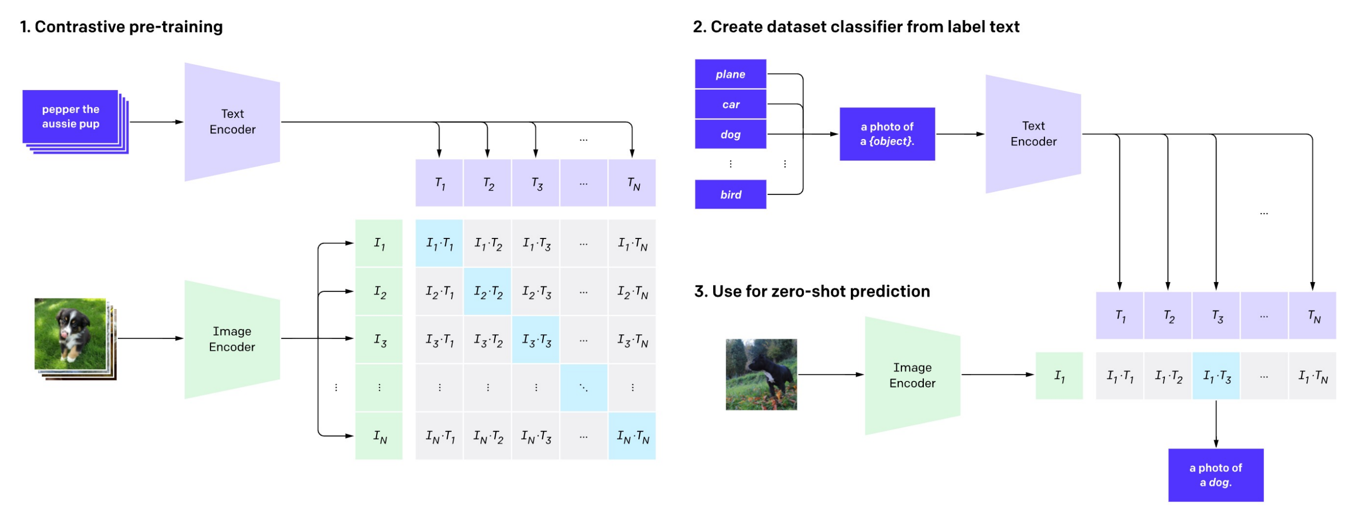
4. Feed Training and prediction into CNN



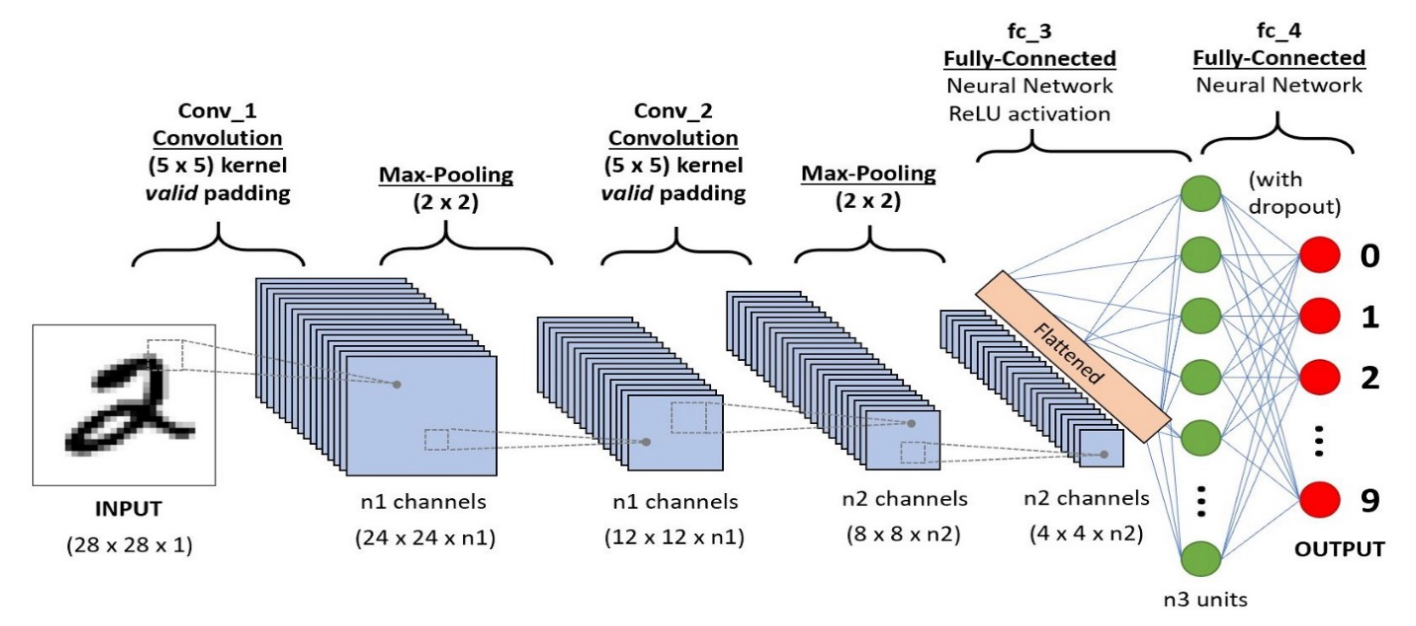
Model Architecture Pipeline

1. Select half chest images and half breast images
2. Stratified Split it into training, and test set

3. Feed training into CLIP

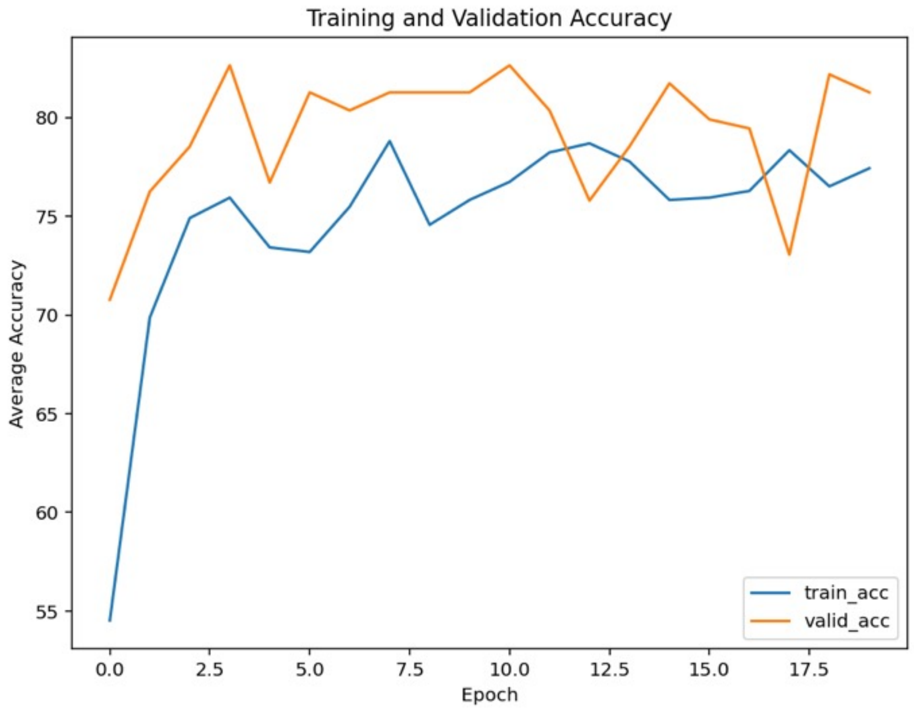


4. Feed Training and prediction into CNN



5. Apply the trained model on the test to compare with results from CLIP only

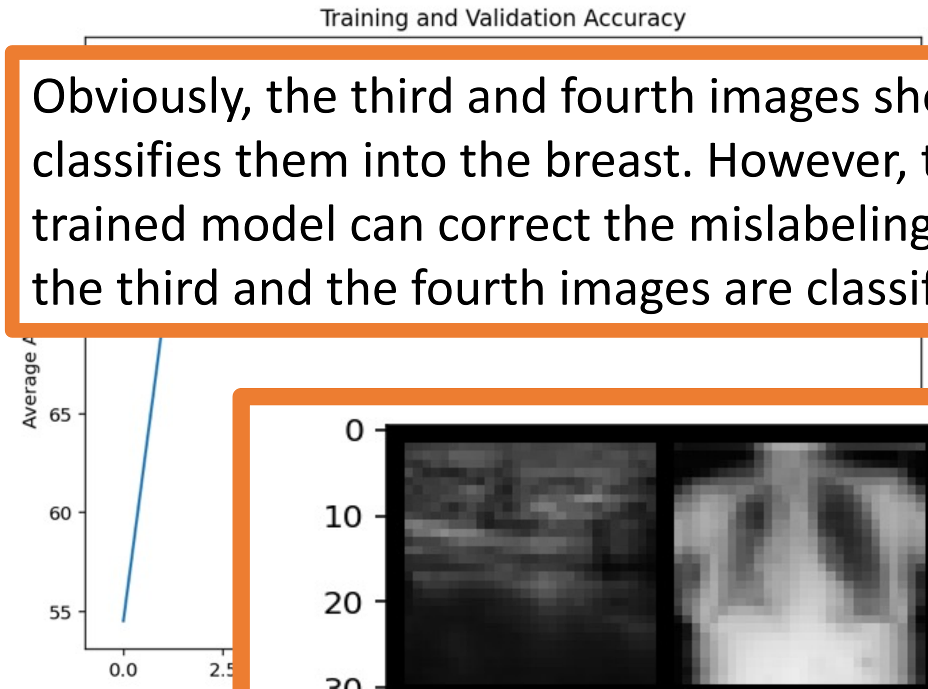
Training and Comparison Results

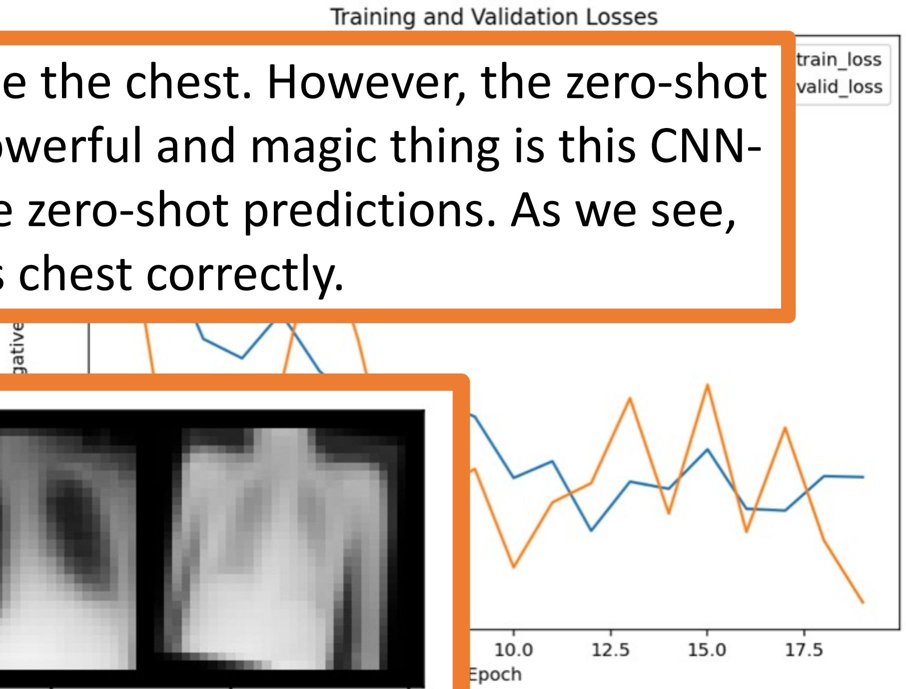


I got 95.72% accuracy on the test set by using trained CNN;

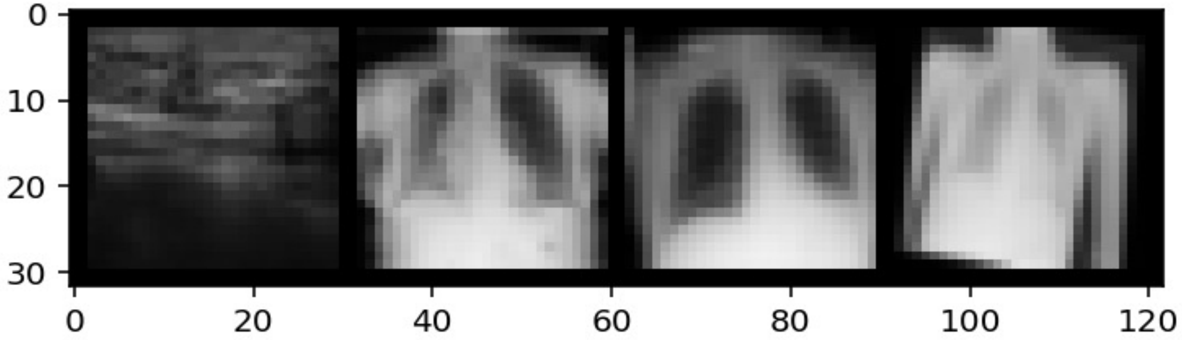
I Got 82.05% accuracy on the test set by using CLIP only

Training and Comparison Results





Obviously, the third and fourth images should be the chest. However, the zero-shot classifies them into the breast. However, the powerful and magic thing is this CNN-trained model can correct the mislabeling in the zero-shot predictions. As we see, the third and the fourth images are classified as chest correctly.

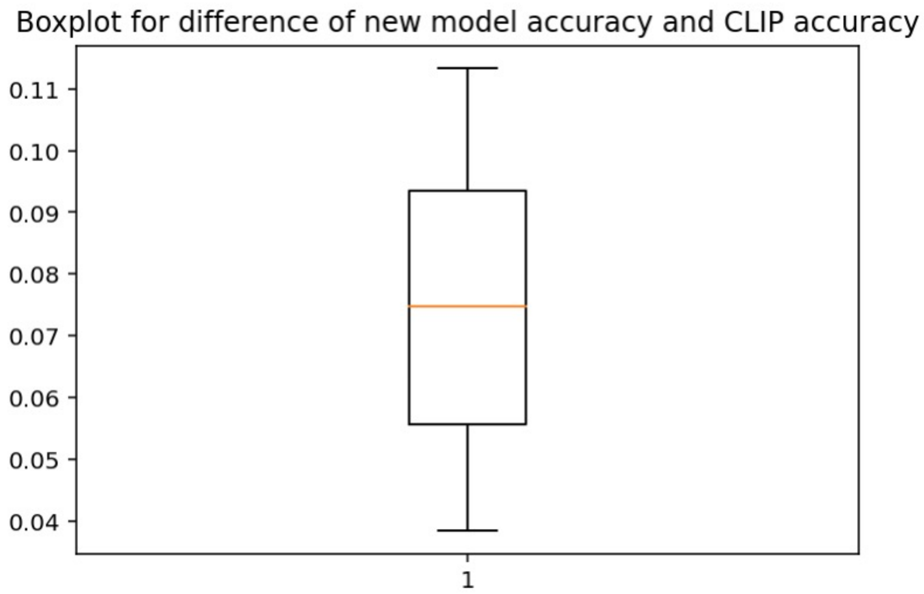


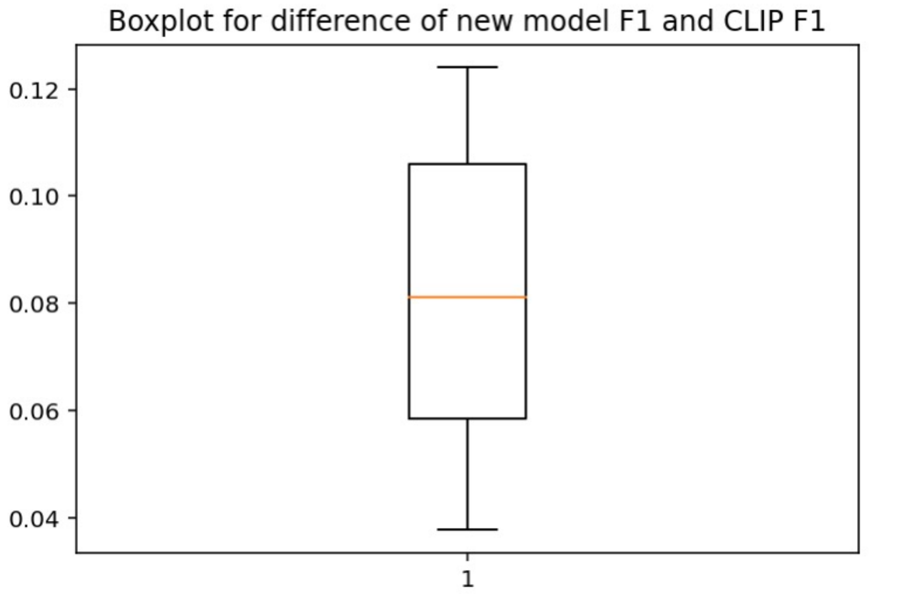
GroundTruth: breast chest breast breast
Prediction: breast chest chest chest

I got 9 ... ng trained CNN;

I Got 82.05% accuracy on the test set by using CLIP only

Run model comparison for 30 times to see generalized improvement





Conclusion

- After these two stages, the trained CNN model can correct mislabel generated by CLIP because CNN is powerful enough to learn the general pattern from majority correct labels and images. The average accuracy improvement is 6%.
- The Two-Stage model can be used to annotate the chest or breast images for production automatically. Then the labeled medical images can be analyzed in the downstream tasks for more research and use cases.

Next Steps

- Employ Monai pre-trained medical image models in the current zero-shot CLIP architecture
- Once the open AI releases and publish its training codes which are used to train their zero-shot learning models between images and texts, I will plan to use those codes to train on all medical dataset in the MedMNIST benchmark dataset and fine-tune them.
- I will plan to test more pairwise datasets, like Blood and Derma, and use different kinds of CNN models

Thanks, Q&A