



# Performance of Biomarkers for CNN Training: A Comparison of Results

Jacob M Ryan,  
Halicioglu Data Science Institute, UC San Diego, La Jolla, CA, USA



UC San Diego Health

## Introduction

Medical imaging requires highly specialized analysis by imaging professionals, a costly and time consuming process. In recent years, there has been a push to use a form of machine learning called Deep Learning to analyze medical images without the need of specialized persons involvement. These models have gained popularity as their effectiveness are proven time and time again. However, the training process for these models requires tens of thousands of medical images and their lab diagnoses.

In an effort to reduce the burden of training the models, this project utilizes biomarkers from blood tests that are representative of pulmonary edema, instead of actual pulmonary edema diagnoses. These blood tests are far cheaper and faster. Therefore, this paper presents models trained on these biomarkers and tests them using unseen and radiologist confirmed images. Additionally, this paper utilizes a lung and heart segmentation model, both to analyze saliency of the previous model and to create new lung-and-heart-focused images to be used for a new model. The goal of this new model is to compare the effectiveness of model prediction on using just the physical aspects of the body we are predicting diagnosis for.

## Hypotheses

1: Deep learning algorithms trained to infer the presence of pulmonary edema from chest radiographs and BNP levels rather than radiologist diagnoses will generalize to unseen data with radiologist diagnoses provided by MIMIC-CXR

2: A deep learning algorithm that emphasizes anatomic structures will perform better than one that is not informed. This will be evaluated by training two algorithms, one provided and tested on lung-and-heart segmentations as additional input channels and one without.

## Data

‘UCSD’ data consists of 16,000 chest radiographs provided by AiDA Lab at UC San Diego Health. ‘MIMIC’ data consists of 22,000 chest radiographs provided by MIT’s MIMIC-CXR study.

## References

Kapishnikov, A., Bolukbasi, T., Viégas, F. and Terry, M., 2019. Xrai: Better attributions through regions. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4948-4957).

## Hypothesis 1: Do Biomarkers Allow for Generalization?

	UCSD Data	MIMIC-CXR
AUROC	0.755	<u>0.877</u>
PRC	0.590	<u>0.799</u>
Accuracy	0.652	<u>0.793</u>

## Hypothesis 2: Does Lung and Heart Segmentation Improve Performance?

	UCSD Data	MIMIC-CXR
AUROC	0.754	<u>0.889</u>
PRC	0.559	<u>0.816</u>
Accuracy	0.616	<u>0.802</u>

## XRAI Analysis

	Input Image	Areas the Model Looked At	Portion of Original Image
True Positive			
True Negative			
False Positive			
False Negative			

## Methods

Each model was built using ResNet 152v2 with ImageNet pretrained weights. Each was trained for 35 epochs using a static learning rate of 1 e-5 and a batch size of 32. Models were analyzed using convectional statistics: Area Under the Receiving Operator Curve (AUROC), Precision Recall Score (PRC), and Accuracy (ACC). During training, models were evaluated every epoch on a withheld validation set originating from the same data as the training set to prevent overfitting.

During testing and evaluation, selected example images from the MIMIC dataset were analyzed on the best performing model using a saliency technique called XRAI. Saliency techniques are a method of visually understanding what parts of each image the model is looking to when it makes a prediction).

There are two experiments being run: comparing efficacy of models trained on biomarkers to those trained on radiologist confirmed data and comparing models trained on the entirety of the X-Ray to those trained on segmented images only. Therefore, each experiment has a model trained only on biomarker data and one trained on radiologist confirmed data. All models are tested solely on radiologist confirmed data.

Post training and testing, example images from both datasets are analyzed using a saliency technique called XRAI. Saliency techniques are method of visually understanding what parts of each image the model is looking to when it makes a prediction.

## Discussion

Hypothesis 1 can be rejected for the moment. Models trained solely on MIMIC (radiologist confirmed) data outperformed models trained on biomarker data. However, the performances were not terrible. There is plenty of room for future experiments to improve upon the work done. Decreasing the cost of training effective medical imaging models allows these models to become more accessible and hopefully benefit more people.

Hypothesis 2 cannot be rejected. There is a non-negligible benefit to utilizing segmentation models that allow classification to focus more easily on anatomically important structures. These segmentation models are increasingly becoming more accurate and more available and with the inclusion of them, results may be better for other use cases.

## Acknowledgments

I would like to thank my mentor Prof. Albert Hsiao for the incredibly valuable insight throughout this project.