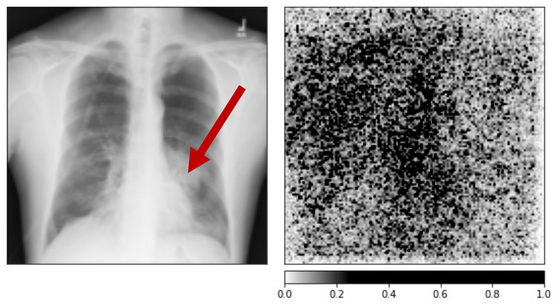
On explainable attention-based deep neural networks trained on radiographic data augmented with diffusion models.

Arnold Caleb Asiimwe; Carl Vondrick, PhD; Mary Salvatore, M.D.

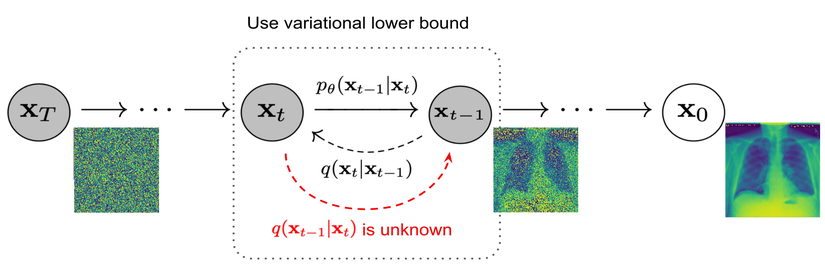
This research presents an **explainable** and **novel neural network architecture (dubbed "ensemble of binary classifiers")** that regularly provides **reliable attributions** for more than 15 classes of cardio-thoracic conditions on chest radiographs augmented with **diffusion models.**

Problem: Multi-label image classification for radiographic data is a strenuous task; results in A.I. models that are not explainable; and performs worse as the number of classes increases.

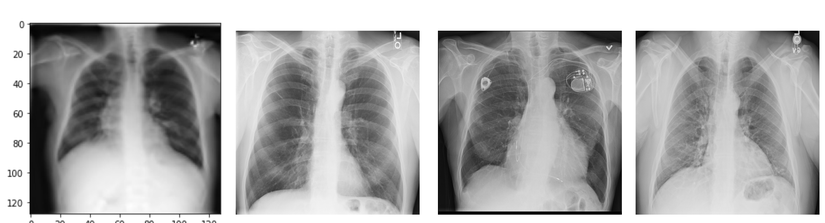


An example of where a 98% accurate multi-label classifier was looking in the wrong part of the radiograph for a mass.

Over 112,000 frontal-view chest radiographs from the National Institutes of Health database are used to train denoising diffusion probabilistic models from which new datasets for each class are generated to augment the training data for the "ensemble of binary classifiers"

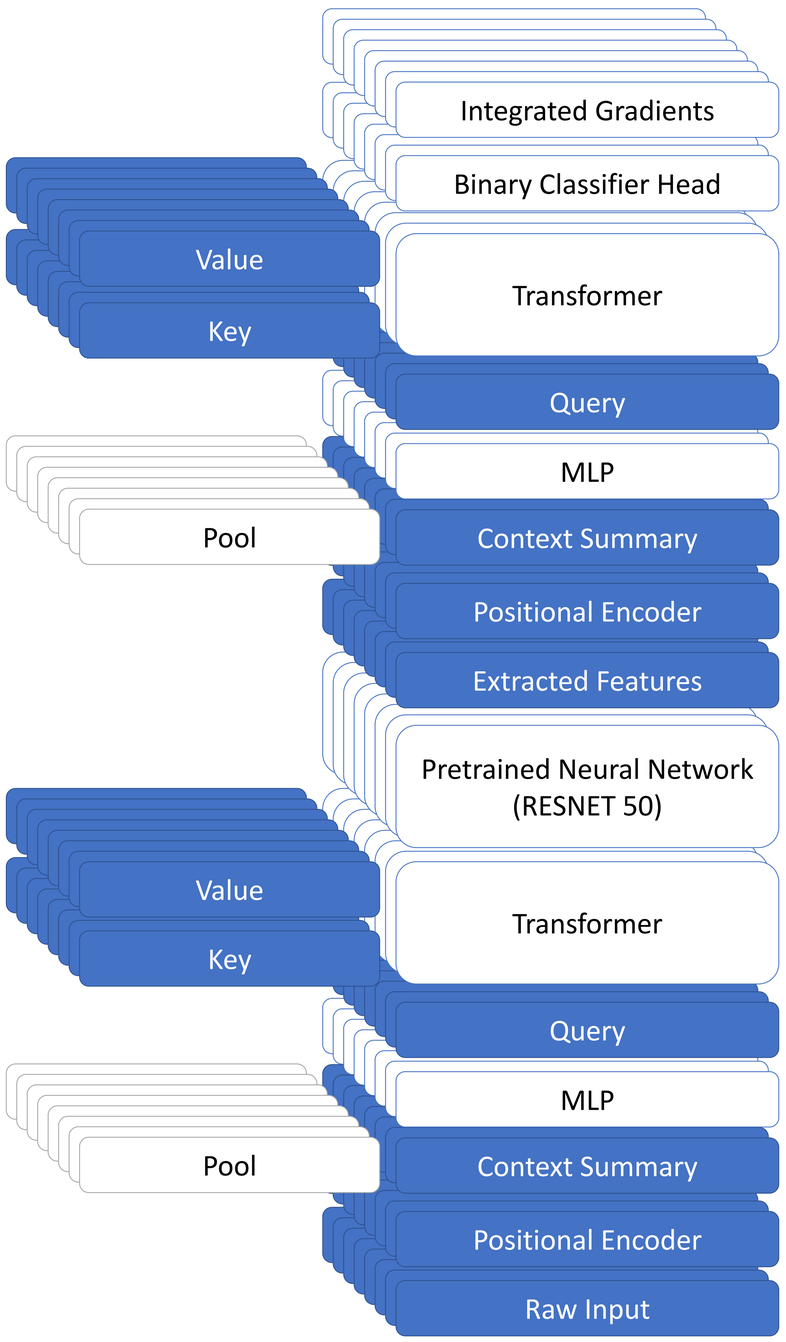


Realistic fake radiographs generated with diffusion models:



"The ensemble of binary classifiers" or EOBC

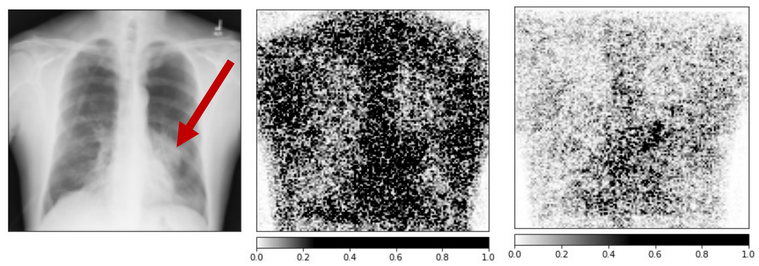
*A series of transformer based binary classifiers trained off a baseline (a normal radiograph) and made interpretable with integrated gradients.



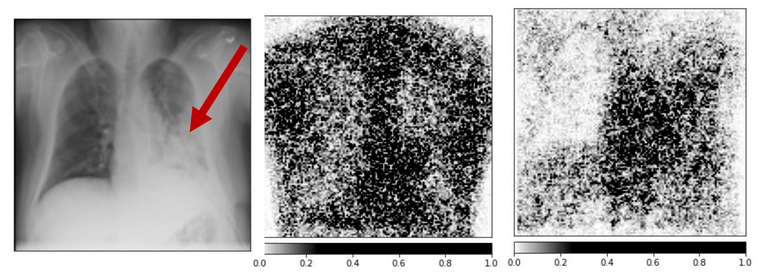
Comparing model explainability of EOBC with conventional multi-label image classifiers.

$$\text{IntegratedGrads}_i(x) ::= \int_{\alpha=0}^{1} \frac{\partial F(\gamma(\alpha))}{\partial \gamma_i(\alpha)} \frac{\partial \gamma_i(\alpha)}{\partial \alpha} \, d\alpha$$

Lung masses:



Pulmonary consolidations



Similar results were obtained for atelectasis, cardiomegaly, edema, effusion, emphysema, fibrosis, Morgagni hernias, infiltrations, nodules, pleural thickenings, pneumonia, and pneumothorax

Conclusion: Training an ensemble of binary classifiers is much easier; yields explainable results and is independent of the number of classes on condition that the right baseline is chosen.

Acknowledgements: Professor Carl Vondrick (Columbia Computer Vision Lab) for being my PI on this project; Dr. Mary Salvatore (Columbia University Irving Medical Center) for teaching me chest radiography; and the I.I. Rabi Scholars Program for funding this research.