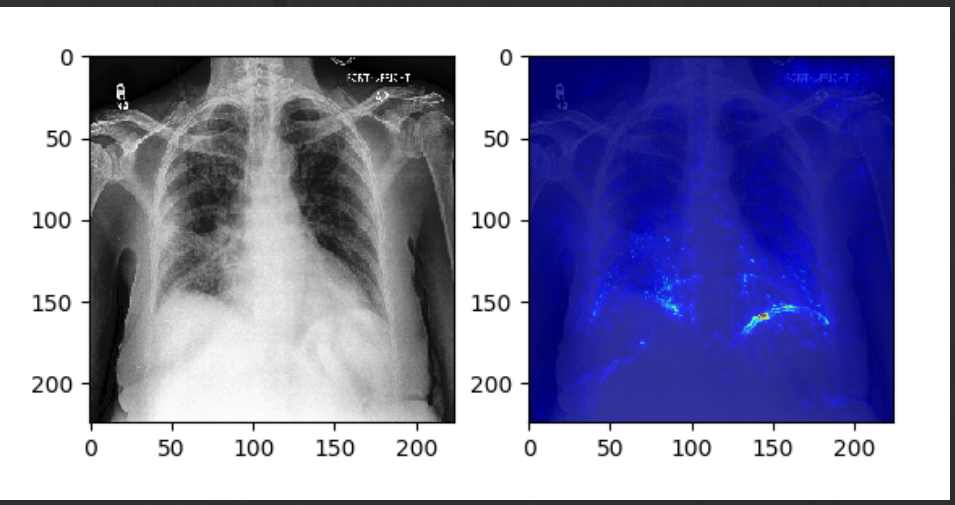
Senior Thesis

Tyler Lian

Assessing Bayesian convolutional neural networks in the clinic

Soon after the advent of the medical X-ray came the first radiologists, doctors to read, label, and interpret the growing range of medical images now essential for modern diagnosis and treatment. It would seem today that computational algorithms are ready to take up the mantle. Researchers have shown that machine learning algorithms are able to "learn" from inputted "training" images to perform as well if not better than human radiologists on important tasks, including the classification of medical images as normal or pathological. Despite these successes, researchers hoping to introduce their machine learning algorithms into clinical practice need to show that their models can robustly hold up to many different dimensions of good performance: Do the algorithms behave according to theoretical expectations? Are the algorithms applicable across different contexts? Are they comprehensible to technicians, physicians, and patients? These questions are often overlooked in medical applications of machine learning.



The empirical convergence, generalizability, and interpretability of Monte Carlo Dropout, a popular implementation of Bayesian convolutional neural networks introduced in Gal and Ghahramani (2016), are assessed in a pneumonia-screening task. In general, Bayesian convolutional neural networks promise to bring the transparency and mathematical rigor that their predecessors have been known to lack, by placing the standard convolutional neural network in a proper probabilistic framework: each model parameter is represented with an updateable distribution rather than point estimates. Findings, however, uncovered several serious shortcomings of these models—they failed to demonstrate the concentration of predictive distributions given more training data, true generalization from one hospital site to another, or resilience to site-specific confounding factors—as well as potential opportunities leveraging the model's quantification of uncertainty. This suggests that more comprehensive research should be pursued before deploying these methods in practice.

Duke UNIVERSITY MATH