Generating Explanations for Chest X-Ray and CT Pneumonia Predictions

Project Abstract

Overview

With the spread of COVID-19, significantly more patients have required medical diagnosis to determine whether they are a carrier of the virus. COVID-19 can lead to the development of pneumonia in the lungs, which can be captured in X-Ray and CT scans of the patient's chest. The abundance of X-Ray and CT image data available can be used to develop a computer vision model able to classify whether a medical image exhibits pneumonia or not. Predictions made by these models can increase the confidence of diagnoses made. Furthermore, rather than teaching clinicians about the mathematics behind deep learning and heat maps, we propose to investigate a new generation of explainable artificial intelligence (XAI) methods whose goal is to annotate images exactly as radiologists do to inform other radiologists, clinicians, and interns about findings. The goal of this project is to develop new methods to annotate medical images with explanations of model predictions that are useful to clinicians and radiologists in analyzing these images.

Methods

Large amounts of labeled medical images have been made accessible for use. This data can be used to build and train a model to make accurate classifications of the presence of COVID-19 pneumonia in medical scans. State-of-the-art models such as UNets and VGGs have performed well in learning the complex features hidden in images. In this project, these models will be trained and optimized through transfer learning on the medical images. Transfer learning is a type of machine learning method that uses pre-trained models as a start, rather than a new model. This allows prior-knowledge that the model has learned to be used while learning the features from new medical scans.

As hand-annotating specific regions of pneumonia present in medical scans is often time-consuming and expensive, XAI algorithms are an alternative and can provide additional benefits. A trained computer vision model is often regarded as a black-box, where an image is inputted and a prediction is outputted with no explanation. In this project, we explore various XAI algorithms such as Layerwise-Relevance Propagation (LRP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM). While the mathematics behind these algorithms differ, the goal is to generate a likelihood heatmap on inputted images to identify regions that contribute to model predictions. This can be done by, for example, propagating through the model's weights and identifying important features, or by testing regions of images to understand the model's behavior. These techniques can produce a model that additionally outputs an explanation, which can increase the confidence in the prediction made.

Constrastive versions of these XAI algorithms will be developed as well, which is a mathematical modification of the original algorithms that produce heatmaps in which different classes are discrimnated against. This allows the differences between pneumonia-positive and pneumonia-negative regions in medical scans to be highlighted more clearly. Furthermore, as CT scans are

3-dimensional images, these algorithms will have to be adapted to work with 3-dimensional computer vision models and generate 3-dimensional explanations. All of these explanations will be generated for tested images and analyzed for accuracy.

Intellectual Merit

The proposed project is potentially useful to clinicians analyzing COVID-19 pneumonia present in medical scans and identifying the regions in which the pneumonia appear. This framework not only generates a classification output, produced by a model trained on large amounts of data, but also highlights regions of each image responsible for the prediction made. Many other applications are possible for clinicians and radiologists performing medical diagnosis on patients. The trained model and explainability algorithms can assist

Broader Impacts Of The Proposed Work

An application can be created using the model and XAI algorithms on the back-end to generate explanations for inputted images. A common concern of AI is the lack of trust in a black-box, which can be detrimental when performing medical diagnosis of the deadly COVID-19. However, with classification explanations and analysis by trained clinicians, the efficiency and confidence of the diagnosis process can be improved.