Project Proposal Title: Providing Visual Interpretations in Concept-Bottleneck Models to Assist In Pediatric Appendicitis Diagnosis

Authors:
Jonas Raedler (jraedler@g.harvard.edu)
Ioana Cristescu (ioanacristescu@g.harvard.edu)
Alyssa Mia Taliotis (alyssamiataliotis@fas.harvard.edu)

Data: We will utilize the Regensburg Pediatric Appendicitis dataset, available on the UC Irvine Machine Learning Repository, published in 2023. The dataset comprises 1709 (multi-view) ultrasound images, as well as tabular data with 782 instances and 53 features.

Background and Motivation: Interpretable Machine Learning is becoming ever more relevant in healthcare, as mere predictions are often not sufficient for doctors to make high-stakes decisions. To address this issue, significant research effort is put into making ML models more interpretable so that healthcare practitioners get an insight into what the model considers when making a prediction.

Problem: Marcinkevics et al. contributed to this effort and developed a concept-bottleneck model (CBM) that assists doctors by identifying concepts from multiview ultrasound images that are relevant for making pediatric appendicitis diagnoses. Doctors can use these concepts to make faster decisions.
We aim to add on to this interpretability by providing visual interpretations of these concepts via, e.g., saliency maps. In essence, rather than just providing the existence of concepts, we want to show doctors *where* these concepts exist in the provided images. This would further speed up the diagnosis process, as doctors could immediately double-check the existence of provided concepts in the actual image, allowing them to more confidently - and more efficiently - make high-stakes decisions.

Scope and Methods: Our project would start by implementing the proposed CBM by Marcinkevics et al. (this consists of a CNN-based encoder, followed by a FNN) and by performing various image preprocessing techniques to clean up the dataset (elements such as annotations, etc. need to be removed from the medical image data). We will then explore various image interpretability methods in order to provide visual interpretations of concepts. As this is not a part of the AC209b curriculum, this exploration also fulfills the extra requirement for AC209b students.

Concerns & Limitations:
1. Ethical: The medical data in this dataset is anonymized so patient privacy will not be compromised.
2. Data Quality: The data seems to be of high quality and well-organized. There do, however, seem to be several missing values that we'd need to address. We also need to perform image preprocessing, but the authors proposed an approach for this.

3. Computational: The images are of size 400x400, so training the model shouldn’t be too computationally expensive. Since the model architecture is described in the paper, we also shouldn’t have too many issues re-implementing a model that hasn’t been discussed in class.