TL;DR

The **basic idea** is to:

- Use pre-trained models that are publicly available for each modality of healthcare data - e.g. X-Ray, MRI, clinical notes, etc.
    - For actual Electronic Health Record (EHR) data (which is structured / tabular data) we can use my previous work - lemonpie if a better one is not publicly available.
- And put them together using one or more of the techniques detailed in this document to predict on data in the 'Coherent Dataset' (and a smaller summary here).

The **risk / caveat** is:

- The **Coherent Dataset** is very small or not big enough for meaningful training.
    - Meaning, the models have to perform pretty much entirely relying on their pre-trained weights.
- If they all perform well - it will be a good example of some type of Data-Centric AI - making models perform well on relatively small datasets and of course transfer learning.
- If they do not perform well, at a minimum - we end up creating a pipeline for multimodal healthcare models which are not available in the public / open source domains, so that will be a significant accomplishment.

Details are in the pages that follow (please scroll down)

- I created the original proposal for another project that did not take off.
    - Because access to the dataset (MIMIC IV) could not be secured on time for the project's team members.
- This is the "scope-reduced version" of the original proposal on the publicly available Coherent Dataset.
    - Where scope is reduced, I have used ~~**strikethroughs**~~ to show the difference.

Problem

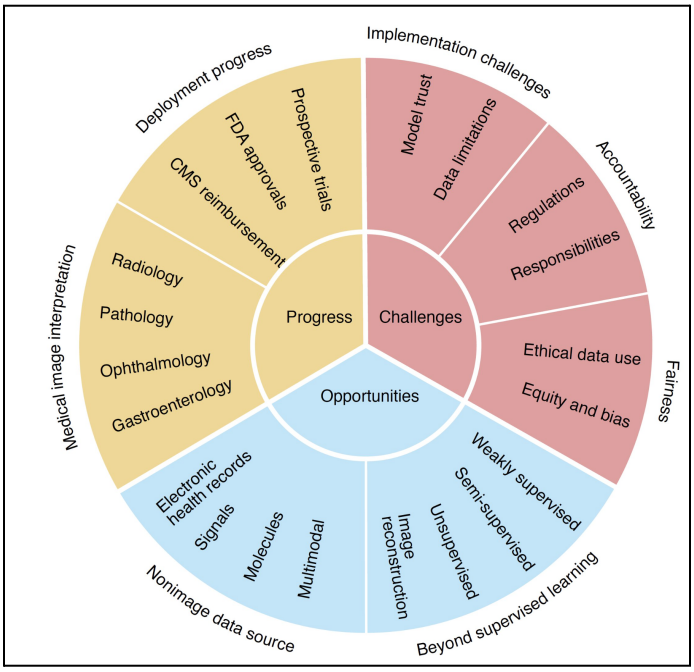
- Physicians (healthcare providers) use data from multiple sources (radiology images, patients' longitudinal history from clinical narrative notes, prior visits, lab reports, etc.) to provide care for patients.
- Yet, the use of such multimodal data in machine learning and specifically deep learning applications in healthcare is not ubiquitous.

This project proposes to develop a **multimodal deep learning model** ~~that surpasses state of the art (SOTA) performance of single-modality models~~ on medical diagnosis tasks (i.e. predicting medical conditions) using Electronic Health Record (EHR) data and Medical Imaging data (e.g. chest X-rays).

Why

Opportunity

- A well regarded paper[1] that reviewed the current state of AI in healthcare, listed models focused on EHR data and multi-modal data to be promising avenues in the near future for advancement of AI in healthcare.



- EHR adoption has significantly improved in the last decade and has reached almost 100% coverage across the US. Yet, not many deep learning models are taking advantage of this rich data source.

[1] Rajpurkar, P., Chen, E., Banerjee, O. *et al.* AI in health and medicine. *Nat Med* 28, 31–38 (2022). https://doi.org/10.1038/s41591-021-01614-0

Multimodal models - based on a recent review paper[2]

- Current SOTA deep learning models in healthcare (e.g. radiology) are dominated by single-modality models.
- Meanwhile advances in other areas like autonomous driving, social media video classification, etc have proven the superior performance of models on multimodal data.
- This paper that did a systematic review of all multimodal efforts in healthcare concluded that:
    - Multimodal models always outperform models based on a single data modality.
    - Deep learning models combining EHR data and images is an area of opportunity (not yet well explored by the research community).

Value-Proposition

- ~~An opportunity to establish a new SOTA for this type of data and model~~.
- Provide a pretrained model for industry and researchers in the field.
- Provide pre-trained embeddings.
- Provide an open source reusable ML pipeline for multimodal healthcare data.

Success

- ~~Success = Surpassing SOTA performance of single-modality models on the same or similar dataset (proposed dataset = MIMIC-CXR + MIMIC-IV)~~
    - ~~That is - surpassing the performance of deep learning models that rely solely on image data (chest X-rays) for medical predictions.~~
- Success = End to end pipeline for predicting using multiple modalities given multimodal healthcare data for a patient.

---

[2] Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine*, *3*, 136. https://doi.org/10.1038/s41746-020-00341-z

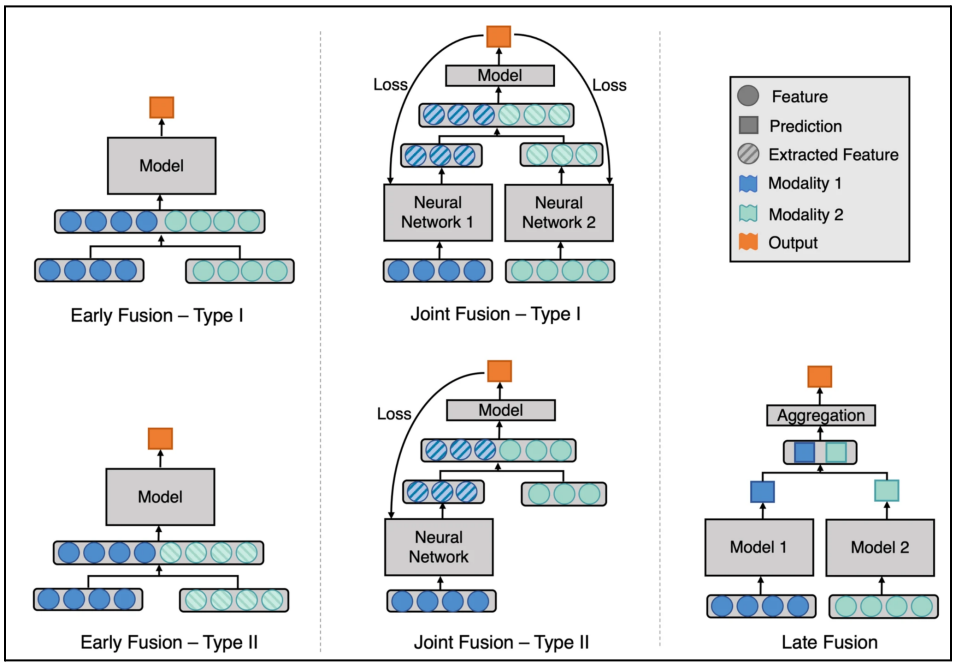
Audience

- Healthcare + AI researchers in academia and industry.
    - Value propositions = An open source ~~and potentially new state-of-the-art~~ model on an open source dataset that:
        - Demonstrates model development on multimodal data.
        - Can be easily benchmarked against and eventually surpassed.
- Health IT teams of hospitals.
    - Pain points = Lack of pre-trained models, embeddings and pipelines.

What

Model(s)

- A deep learning model or comparative models implementing one or all of the multimodal fusion techniques (early, joint and late fusion) described in the paper mentioned above[2].



- Model for image data = CNN (some type of Resnet).
- EHR data (Structured + time series) = LSTM or potentially Graph NN.
- Potentially Clinical Notes = Transformer (BERT).

Dataset

- ~~The relatively new MIMIC-IV and MIMIC-CXR datasets provide the only English-language data (available anywhere in the world) of actual EHR and related radiology images for de-identified patients.~~
- ~~Risk: Timely access to MIMIC for all team members~~
- The Coherent Dataset.

Metrics

- AUROC - the most popular metric in healthcare applications of this type.
- All current papers and results are published and measured based on AUROC.