IN5310 - P3 project proposal: Enhancing Multi-Modal Deep Learning for Opportunistic Risk Assessment in OL3I

Berezin, Ilya & Kumar, Nikhil

October 20, 2023

1 Motivation

Ischemic heart disease (IHD) is a leading cause of death globally. Early prediction and risk assessment are crucial for timely interventions. The OL3I dataset, comprising CT images and tabular medical data, provides a unique opportunity for opportunistic risk assessment. While the initial study utilizing OL3I employed deep learning, there remains potential for enhancement, particularly in the areas of multimodal integration and model explainability.

2 Baseline Method

The original OL3I study combined convolutional neural networks (CNNs) for CT image analysis and fully connected networks for tabular data. The predictions were then interpreted using SHAP (SHapley Additive exPlanations). This approach serves as our baseline.

3 Datasets

The Stanford OL3I dataset will be the primary data source. It contains 8,139 axial CT slices at the L3 level and 422 features from medical records. Labels indicate individuals diagnosed with IHD 1 or 5 years post-scan.

4 Evaluation Metrics and Methods

To ensure our model's robustness and its alignment with the original study, we will employ the following evaluation metrics:

- **Area Under the Receiver Operating Characteristic Curve (AUC-ROC)**: A primary metric that evaluates the model's ability to discriminate between positive and negative classes.
- **Sensitivity, Specificity, and Accuracy**: To provide a holistic view of the model's performance.
- **Positive Predictive Value (PPV)**: To gauge the model's precision in predicting positive instances.
- **Brier Score**: To assess the calibration of the predicted probabilities, ensuring the reliability of the model's confidence in its predictions.

To compare the multi-modal model's performance to the uni-modal case, we will train separate models on each data modality and evaluate them using the mentioned metrics. A direct comparison will highlights the benefits of multi-modal integration.

5 Potential Enhancements (might be still limited)

- **Attention Mechanisms and Multi-Modal Architectures**: We will try to integrate attention mechanisms to allow the model to focus on relevant parts of the CT image and significant features in the tabular data. This can potentially enhance the predictive accuracy. Exploring transformers architectures that have shown promise in multi-modal tasks, could help us in better fusion of image and tabular data.
- **Enhanced Explainability**: Beyond SHAP, we propose integrating techniques like Grad-CAM for visual explanations on CT images, providing clinicians with more intuitive insights into model decisions.
- **Transfer Learning**: We will use models pre-trained on larger, similar datasets to improve performance, especially given the challenges of medical data scarcity.

6 Proposed work plan

The work would be splitted. Time frame of remaining weeks should be sufficient, where the first week is already utilized for dataset and methods related research and the rest will be used for the project implementation. Last week to be reserved exclusively for the report writing.