Paper/ Slide Name
SHAP based evaluation of Model Debugging and learning transferability

Current Research Status and Work Done Since Last report

Model Training and Fine-Tuning

- I have tried several pre-trained architectures, including InceptionV3, ResNet101, and VGG19, as the base models. These models were fine-tuned using brain cancer and then tested with SHAP. I saw the best results using Inception V3.
- The fine-tuning process involved unfreezing the last few layers of each model to allow for domain-specific training. For InceptionV3, the last three layers were unfrozen; for ResNet101, the last seven layers; and for VGG19, the last eight layers. Again, like mentioned earlier, I got the best results when I used InceptionV3.
- The models were optimized using Adam optimizers with hyperparameters tuned for the specific dataset. Callbacks such as ReduceLROnPlateau and EarlyStopping were employed to enhance training efficiency and prevent overfitting.

SHAP values were computed for the InceptionV3 model to identify key features influencing the model's predictions. This involved analyzing which aspects of the lung scans were critical for distinguishing between benign, malignant, and normal cases. Initial findings show differences in SHAP values between the lung and brain datasets. This suggests that while some common features may be consistently important, others may not transfer well, showing the need for domain-specific models.

Worked on the design and layout of the poster. Focused on making the content easy to follow and understand and have something to work off of where I don't have to read off the poster.

Hurdles and Obstacles

Challenges with SHAP Implementation

- The biggest challenge has been working with SHAP. While there are different types of SHAP explainers like kernel and deep explainers, i have only been able to use the .explainer method. This is because I have run into technical problems with the other options, limiting our analysis.
- Using Google Colab has been limiting my work. It doesn't have enough processing power and memory for the large datasets and complex models we're working with. This has caused issues like incomplete analyses and slowed down progress. Google Colab, while convenient, doesn't provide enough resources for our needs. The limitations in runtime and hardware capabilities have been a significant bottleneck. Having access to a dedicated server with more power and memory would greatly improve our ability to perform detailed analyses.

Three Boxes (Green-Yellow-Red) Framework

1. Green Box: Based on Literature/Initial Work

- Foundation and Background: My research builds on existing studies that use SHAP to explain how deep learning models make decisions, particularly in medical image analysis. These foundational studies have shown that SHAP can help identify important features in medical

images and aid in understanding model predictions. I'm using this established knowledge as a starting point for my work.

2. Yellow Box: Research Filling the Gap

- While SHAP has been widely used, there hasn't been much research on how well these explanations transfer between different datasets. My work focuses on testing whether SHAP explanations from a model trained on lung scans can be applied to brain scans. This helps me understand how consistent and reliable SHAP is when used across different types of medical data.
- By exploring SHAP transferability, I'm looking at a new area in model interpretation. I'm examining how feature importance changes between domains, which could lead to more general or customized explanation methods. This work is important because it highlights the challenges of using a single tool for different medical conditions, suggesting that more specialized explanation techniques might be needed.

3. Red Box: The Bigger Picture

- My broader goal is to make AI models in healthcare more reliable and trustworthy. By improving model interpretability, particularly with tools like SHAP, I want to help clinicians make better decisions. This research is part of a larger effort to create AI systems that are not only accurate but also transparent and easy to understand.
- Understanding how SHAP explanations work across different types of medical data can help create models that are more versatile and useful in various clinical situations. This research might also influence future guidelines for using AI in healthcare, ensuring that models provide consistent and reliable explanations no matter the medical context.
- My findings could lead to more studies on cross-domain model interpretation and the development of new or improved explanation tools. Ultimately, I aim to connect technical advancements in AI with practical applications in healthcare, improving patient care and outcomes.

Evaluate My work

I will compare SHAP values across different datasets (lung and brain scans) to evaluate the consistency of feature importance. This involves analyzing how well the rankings of important features match between the two domains. A high degree of similarity would indicate good transferability of SHAP explanations.I will assess the performance of the InceptionV3 model on both the original (lung scans) and new (brain scans) datasets. Metrics such as accuracy, precision, recall, and F1 score will be used to measure the model's effectiveness in making accurate predictions across different types of data.By visualizing SHAP values using plots like summary plots and force plots, I can qualitatively assess the explanations provided. This visual inspection will help in identifying patterns that may not be shown through quantitative numbers alone.

Research Remaining // Steps

- Make the code and data available to others, ensuring they can access everything needed to reproduce the study. Making sure everything is clear.
- Finalize the research paper and poster presentation. Make sure to clearly explain what was done, the results, and why it matters.
- Look closely at where the model makes mistakes and work on reducing these errors.

- Wrap up the analysis of SHAP values between lung and brain scan datasets.