



# Summer Internship 2024

<<Interpret-CXR >>



*Team members:*

*Faculty-in-charge:* Dr. Sowmya Kamath S.

Department of Information Technology  
National Institute of Technology Karnataka, Surathkal

**June 7, 2024**

# Language Model Eval Metrics

Findings Language Model

**Perplexity:** 1.9489982488623694

**BLEU:** 0.3722908950014124

Impressions Language Model

**Perplexity:** 1.9978719323448744

**BLEU:** 0.7330372239916698

Both the Models are trained to 3 Epochs going over that reduced the score

## Setup a joint model

Integrating a **Swin Transformer** as a vision encoder into the pipeline that includes a language model. Goal is to map the image features extracted by the Swin Transformer into the same embedding space used by the language model.

Preprocessed the images as needed by Swin Transformer

**Input Image -> [Swin Transformer] -> 768-dimensional feature vector**

**768-dimensional feature vector -> [Linear Layer] -> Projected 768-dimensional vector**

**Projected 768-dimensional vector -> [Tanh Activation] -> Final 768-dimensional embedding**

**Final 768-dimensional embedding -> [Language Model] -> Downstream tasks (e.g., text generation, classification)**

**THANK YOU**