Problem description

Problem statement

Interpreting and summarizing medical images has become more and more important, and so the need of automating it. This is because experts might differ sometimes on their opinions and they require a lot of preparation to have a good judgement, thus, having a reliable system that can always understand and explain medical imaging is vital and will help to avoid several of the bottlenecks that health systems have to deal with nowadays.

This project is intended to give a step towards that goal, through the participation in the 7th edition of the Caption Task in the Challenge called *ImageCLEF 2023.* In this challenge, descriptive and coherent captions are desired to be generated based on a given image.

Hypothesis

In order to solve this problem, a recent architecture will be used, it's called *Transformers.*

Among other things, a comparison of a trained-from-scratch model vs re-training a given model will be explored.

Metrics

The metrics used to evaluate this model in the challenge will be:

- BERTScore
- ROUGE

And other metrics such as METEOR, CIDEr, and BLEU wil be reported as well.

Dataset

An updated and extended version of the *Radiology Objects in Contexts (ROCO)* will be used. It contains images, their captions and associated UMLS terms.

The size of the dataset is:

- Training set: 60,918 radiology images.
- Validation set: 10,437 radiology images.
- Test set: 10,437 radiology images.