During the initial presentation, our deliverables changed. For our second checkpoint, we've prepared one page of written rapport and a jupyter notebook with example code and example results. We had to create an explanation for vision transformers using existing repositories. If we followed the original project's description we would use class-agnostic methods as our explanation. We chose to explain vision transformers with the existing class-specific method which gives us the opportunity to highlight the activations of neurons across classes.

What was done?

We trained a vision transformer to solve the task of lung classification of people who suffered during the pandemic of COVID-19. In order to explain it we used the implementation of LRP explanation from github called Transformer-Explainability authored by Hila Chefer, Shir Gur and Lior Wolf from the University of Tel Aviv published at The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR). Example code of this paper explained our vision transformer on ImageNet and it could highlight what part of the image excites the output neuron. Our dataset had 3 possible classification classes:

- COVID-19,
- Non-Covid - other lung infections (Viral or Bacterial Pneumonia),
- Normal - lungs of a healthy person.

Since our dataset was balanced we used accuracy as our main metric, we achieved:

- Train dataset - 88%,
- Validation dataset - 77%
- Test dataset - 75%

What are the difficulties?

Transformer training is difficult. SGD optimizer wasn't effective enough for this type of machine learning model, but Adam yielded better results. We might try to use different optimizers and learning rate schedulers to further increase the accuracy of our prediction, thus improving the quality of our explanations. Another problematic factor is the lack of sufficient computing resources while university-provided clusters had heavy traffic when we wanted to train our transformer.
We also encountered some artifacts in a major part of the dataset where our explanations highlight a region in the corner of the image, often outside of the patient body.

What will be done next?

- Train model on a higher resolution (currently 224x224) and compare results,
- Hyperparameter tuning,
- (Optional) Train model on melanoma dataset,
- (Optional) Compare results with resnet/vgg