AI in Medicine I

Practical exercise 4

Submitted to the
Department of Informatics
I31 - AI in Medicine and Healthcare
I32 - Computational Imaging in AI and Medicine
of TUM

by

Argudo, Mateo (03717216)
Bolaños, Daniela (03765336)
Frey, Daniel (03632203)
Javadov, Aydin (03749463)
Otto, Julia (03698083)

January 12, 2024

1 Interpretability and Explainability

1.1 Using CAM-based methods

For this exercise we were tasked to compare different CAM-based methods. CAM based methods are used by the user to understand which areas in an image are relevant for the model to select a class (in case of a classification problem). In this solution five CAM-based methods are analysed, namely: GradCAM, GradCAM++, EigenGradCAM, AblationCAM, RandomCAM. To compare the performance of the CAM-based methods two models are chosen: Resnet50 and Resnet101. This allows to take a look at generalizability and transferability. The models are each trained for ten epochs. Similarly, two data sets are chosen of similar structure (input channels and classification): BloodMNIST and PathMNIST. The implementation allows for a variable number of trainable layers, which adds an additional parameter for tuning and improving the models (two trainable layers are chosen during inference). Since choosing the very last layer of a network as the target layer for CAM methods led to inconclusive results, the second-to-last layer was picked instead. For visualization of the results, images were upscaled to 100 pixels $\times$ 100 pixels (width $\times$ height).

For both data sets the choice of the model does make a difference in the distribution of the areas of interest of the model and consequently user readability (see figure 1). Whilst for Resnet50 often a broader area is highlighted, a more defined area is highlighted evaluating Resnet101. It can be concluded, that the more complex model in this case allows for better intepretability. Whilst the choice of different data set does not show a large difference in readability, the CAM-based methods, when viewing the higher interpretability of the Resnet101 model, do. Especially, GradCAM and AblationCAM seem to give the most defined area. The areas also

fit well to objects in the original image. It is also interesting to observe that sometimes the areas of interest are inverted between different methods (see GradCAM, GradCAM++ for Resnet101, PathMNIST). For this reason the methods would be ranked according to the following list, with the first method mentioned containing the most distinguishable visual information for an untrained user:

1. AblationCAM
2. GradCAM
3. EigenGRadCAM
4. GradCAM++
5. RandomCAM

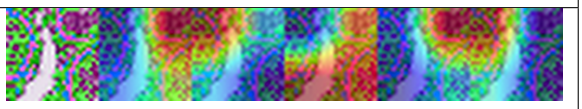
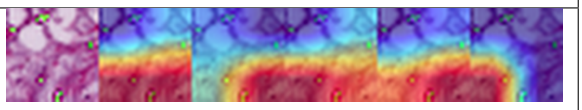
Model	Data Set	Accuracy	Areas of interest
resnet50	bloodmnist	81.1%	
resnet101	bloodmnist	83.4%	
resnet101	pathmnist	86.7%	
resnet50	pathmnist	87.0%	

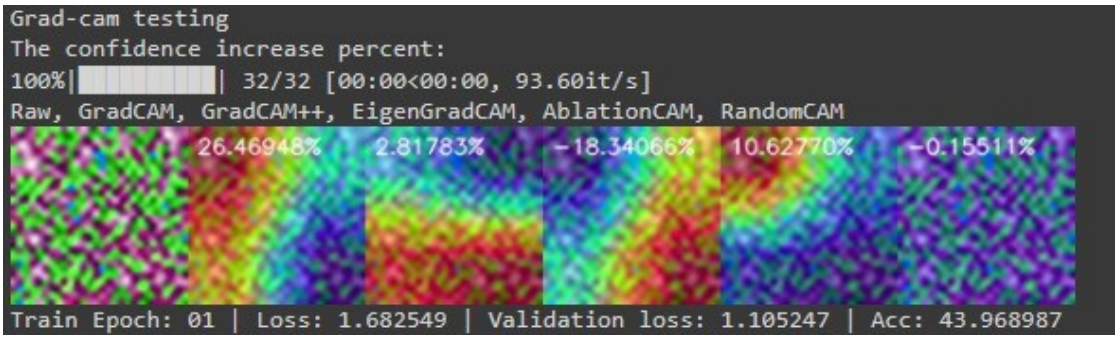
Figure 1: Figures showing the CAM-based areas of interest evaluated. From left to right: Raw, GradCAM, GradCAM++, EigenGradCAM, AblationCAM, RandomCAM.

1.2 Analysing the feature selection

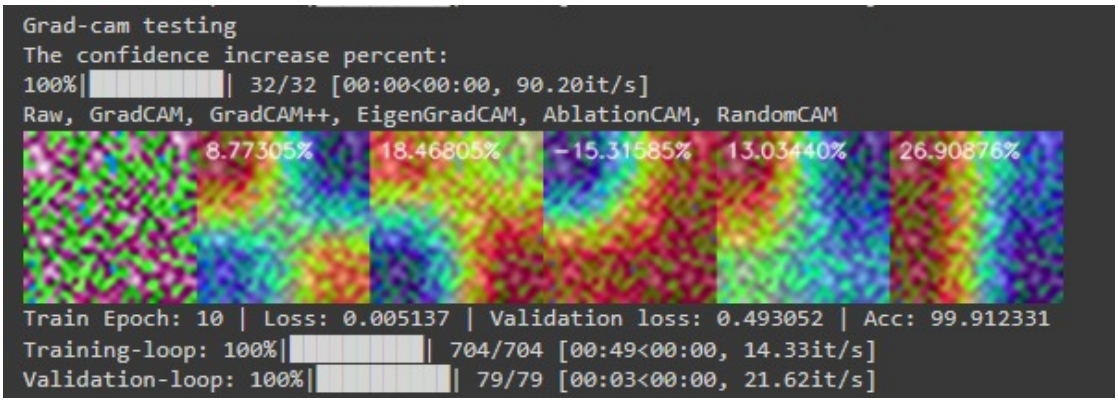
1.2.1 PathMNIST

According with the Fig. 2 we can see the low region overlap between the performed segmentation of the histological slides of the untrained model and the areas with the high activation (heat maps). Nevertheless, after training the model for 10 epochs, we can observe a better overlap between regions of high activation and the histological cancer cells, even we could detect two region activation maps in GradCAM and GradCAM++, that we could not identify in our untrained model, as well as in RandomCAM, the presence of high activation due to this overlap.

In addition, the large confidence change is another sign that the model started to learn useful features, and the accuracy reaches until close to 100% which could means a good sign for training the model and learning these features, without increasing too much the validation loss, to avoid overfitting of our model. Also, as we did the performed interpolation due to the use of ROAD metrics, our image contains a bigger confidence due to the pixels with high activation in some regions, e.g., in EigenGradCAM and RandomCAM methods.



0 epochs



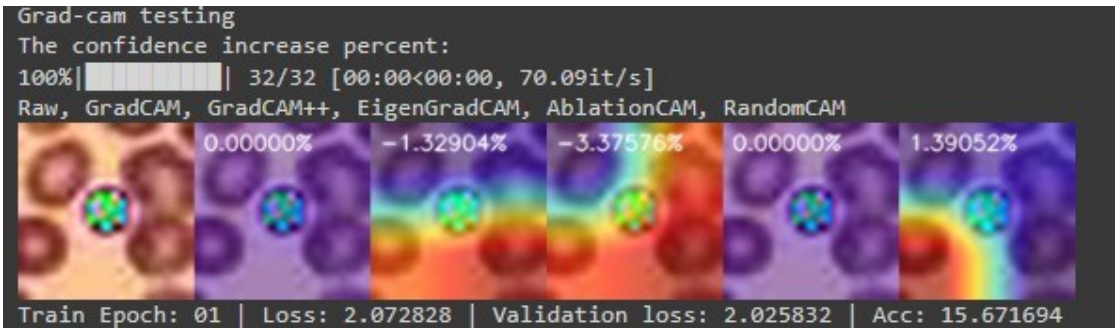
10 epochs

Figure 2: CAM-based evaluation of untrained and trained model using the PathMNIST dataset.

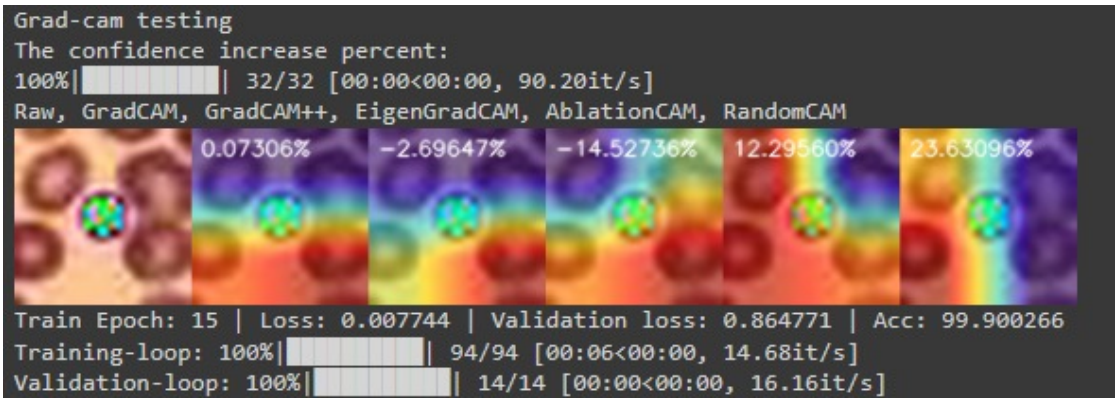
1.2.2 BloodMNIST

In this dataset, we can observe a bigger difference between the blood cells segmentation and the high activation maps that recognized for our untrained model. We could observe the accuracy and compare it for both untrained and trained model with 15 epochs, and we find a big difference on it, which means our model already is good enough with the learned features until that specific number of epochs, and our validation loss is not increasing anymore, which leads to avoid overfitting of our model. In terms of the confidence, we can see a big change between untrained and trained model (e.g., RandomCAM), it could means that our trained model

with 15 epochs identify more cells segmentation regions with an improved overlap within the heat activation maps. Basically, the untrained model has a random activation map, instead of the trained model with 15 epochs, which presents a good prediction of the blood cells segmentation.



0 epochs



15 epochs

Figure 3: CAM-based evaluation of untrained and trained model using the BloodMNIST dataset.

2 Monte-Carlo Dropout to estimate uncertainty

2.1 Q1: Loading the Pretrained Model

As stated we loaded the Resnet18 Model with the pretrained weights, added the Dropout layer with $p = 25\%$, changed the output layer to return 7 classes, and

freezed all the layers but the fully connected layer.

2.2 Q2: Runtime Error Correction

We corrected the dimensions of the labels from $[[y_0], [y_1], ..., [y_n]] \xrightarrow{to} [y_0, y_1, ..., y_n]$ using the squeeze operation from torch library.

2.3 Q3: Calculating the Accuracy

We calculated the accuracy which is 74%.

2.4 Q4: Examining model uncertainty

To examine the results we first run an MC simulation for the classifier scores but the model stayed in evaluation mode, thus dropout layer remained inactive. After switching it back to training mode the dropout layer was active again. Hence we could observe how uncertain are scores for given samples. For most of them scoring was quite consistent. However, one of them was uncertain between different classes. The result is shown in figure 4.

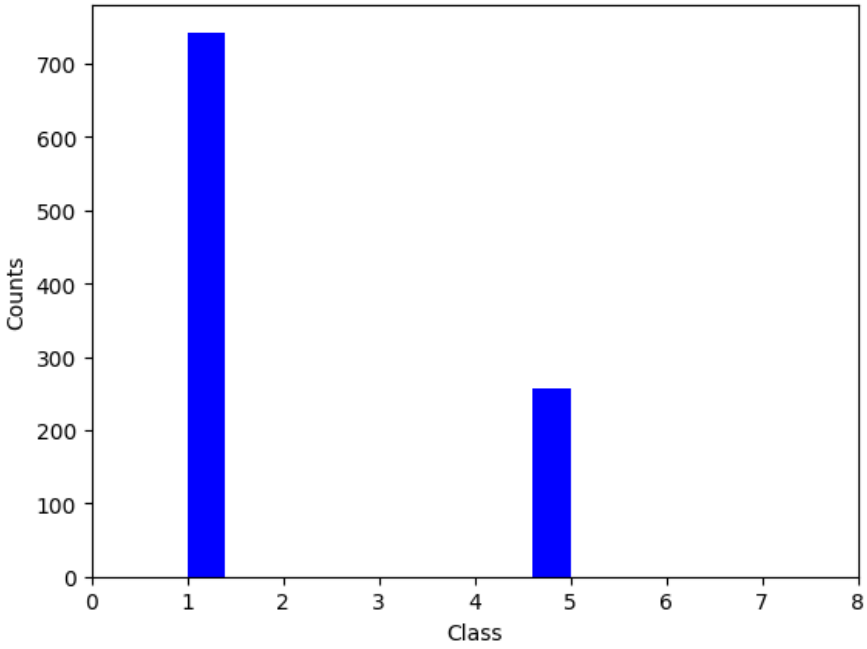


Figure 4: Results obtained from plotting the predictions of the model. The model was uncertain between to classes.