A blue and black logo

Description automatically generated

Trustworthy AI

AI in Medicine I: Practical Exercise 4

# **Authors:**

|  |  |
| --- | --- |
| **Name** | **ID** |
| Mohammed Elbushnaq | 03786474 |
| Muhammed Elmasry | 03786399 |
| Evangelos Fragkiadakis | 03786480 |
| Nada Elsherbeny | 03788337 |

Note:

**Evangelos Fragkiadakis’s** ID is **03786480**. However, it was written as **03784680** in the first two assignments. We apologize for this mistake and hope that it didn’t cause any issues with you.

# Task 1

|  |  |
| --- | --- |
| Dataset | bloodmnist |
| Labels | {'0': 'basophil', '1': 'eosinophil', '2': 'erythroblast', '3': 'immature granulocytes(myelocytes, metamyelocytes and promyelocytes)', '4': 'lymphocyte', '5': 'monocyte', '6': 'neutrophil', '7': 'platelet'} |
| model | resnet50 |
| Training Accuracy | 85.50% |
| Test Accuracy | 83.87% |
| CAM-based Methods used | GradCAM++, EigenGradCAM and RandomCAM |

A close up of a circle

Description automatically generatedA blurry image of a purple circle

Description automatically generated

Figure 1: training and test samples and their labels

(a) Using CAM-based methods

**A purple circle with text

Description automatically generated**We can see samples from the dataset in Fig. 1, the sample we will use in visualization in Fig. 2,

As shown in Fig. 3, we can see that the best one is the EigenGradCAM.

Figure Sample used for Visualization

The GradCAM++ also looks good, but EigenGradCAM is much better, and the randomCAM is the worst, as also shown in Fig. 3.

EigenGradCAM is better than GradCAM++ and RandomCAM because it provides more accurate, sensitive, and specific visual explanations of what a deep learning model focuses on in an image, as clearly shown in Fig. 3. It's also more effective in highlighting crucial features and is generally more robust to noise. Furthermore, the EigenGradCAM focuses mainly on our Region of Interest, while the GradCAM++ takes extra information from the area surrounding our ROI, and the RandomCAM generates heatmaps based on randomness, making it the least effective one, leading to low specificity, interpretability, and consistency in explaining model decisions.

A screenshot of a graph

Description automatically generated

Figure 3 Visualization of differen CAM-based methods in the sample

# b) Analysing the feature selection

A screenshot of a graph

Description automatically generatedEpoch 1:

Figure 4 The first epoch of the training

### Epoch 2:

A screenshot of a graph

Description automatically generated

Figure 5 The second epoch of the training

### Epoch 3:

A screenshot of a graph

Description automatically generated

Figure 6 The third epoch of the training

As shown in Fig. 4, Fig. 5, and Fig. 6, we can clearly see that the model is converging.

The heatmaps across the three epochs show that the GradCAM++ and EigenGradCAM methods trend towards more focused and intensified regions of interest, indicating that the model is learning to better identify the features necessary for making predictions. GradCAM++ shows a progressive refinement from a broader spread towards a narrower area of focus, while EigenGradCAM maintains a consistent focus with increased clarity over time. This suggests that the features selected for explanations by the model are becoming more representative of the underlying patterns it is learning to recognize.

RandomCAM, on the other hand, does not show any meaningful progression, as it is based on random factors and thus remains scattered and non-informative throughout the training process. Considering the progression in the heatmaps, EigenGradCAM provides stable and precise feature selection by the third epoch, potentially indicating a point where the model's understanding of essential features has matured.

To ascertain the optimal stopping point for training, one should also factor in quantitative performance measures alongside these qualitative insights to ensure the model not only focuses on the correct features but also performs optimally according to the chosen metrics. Moreover, on epoch no. 12, EigenGradCAM gives a ROAD value of 0.10114. The exact value is -0.21816 on epoch no. 15, which indicates that the model is overfitting, and the feature map activation is not providing any further value.

In conclusion, we can clearly see that 12 epochs are enough for the training process not to encounter any overfitting.