8P361 Project Imaging - BIA Group 4 | Assignment 3
Myrthe van den Berg, Rick van Bergen, Nils Bodenstaff, Cris Claessens and Jeanine van Ingen

Exercise 1

The PatchCAMELYON challenge on Kaggle uses the area under the ROC curve as an evaluation measure. Describe the concept of ROC curve analysis and the area under the ROC curve (AUC) as an evaluation measure. Then, using methods from the sklearn Python toolbox (the required functions are already imported), perform ROC curve analysis and computation of the AUC for your model.

Note that you will have to do this analysis on the validation set (since you do not have access to the ground truth for the test set).

The ROC (receiver operating characteristics) curve is a probability curve with TPR (True Positive Rate) on the y-axis and FPR (False Positive Rate) on the x-axis. The TPR is the same as the sensitivity which can be calculated with equation 1. The FPR is the same as 1-specificity and can be calculated with equation 2. In the case of the PatchCAMELYON challenge, a TPR of value 1 means that all the samples containing metastases are classified as showing metastasis. On the other hand, an FPR of value 1 means that every single sample in the PatchCAMELYON dataset that was not showing metastasis is incorrectly classified as showing metastasis [1].

$$True\ Positive\ Rate = \frac{True\ Positives}{True\ Positives + False\ Negatives} \qquad \textit{(Equation 1)}$$

$$False\ Positive\ Rate = \frac{False\ Positives}{False\ Positives + True\ Negatives} \qquad \textit{(Equation 2)}$$

The Area Under the Curve (AUC) represents the degree of separability and tells how much a model is capable of distinguishing between models. The higher the AUC, the better the predictions the model is returning. With an AUC of 1, all predictions are correct, so samples belonging to the class 0 are classified as 0 and vice versa for class 1. With an AUC of 0, all predictions are the opposite of the actual value, so samples belonging to the class 0 are classified to 1 and vice versa. When the ROC is a straight diagonal line, in which for every point on the line the TPR equals the FPR, the AUC is 0.5. In that case, the model has no class separation capacity at all [2]. AUC is mainly used and preferred over accuracy for binary classifications, as it deals much better with skewed distributions. It selects models with false and true positive rates that are significantly above random chance. For accuracy, this is certainly not guaranteed [3].

In figure 1 the ROC analysis of the given model is shown. Since the AUC is 0.93, this means that the model predicts quite adequately already, but still needs some optimization to make even better predictions.

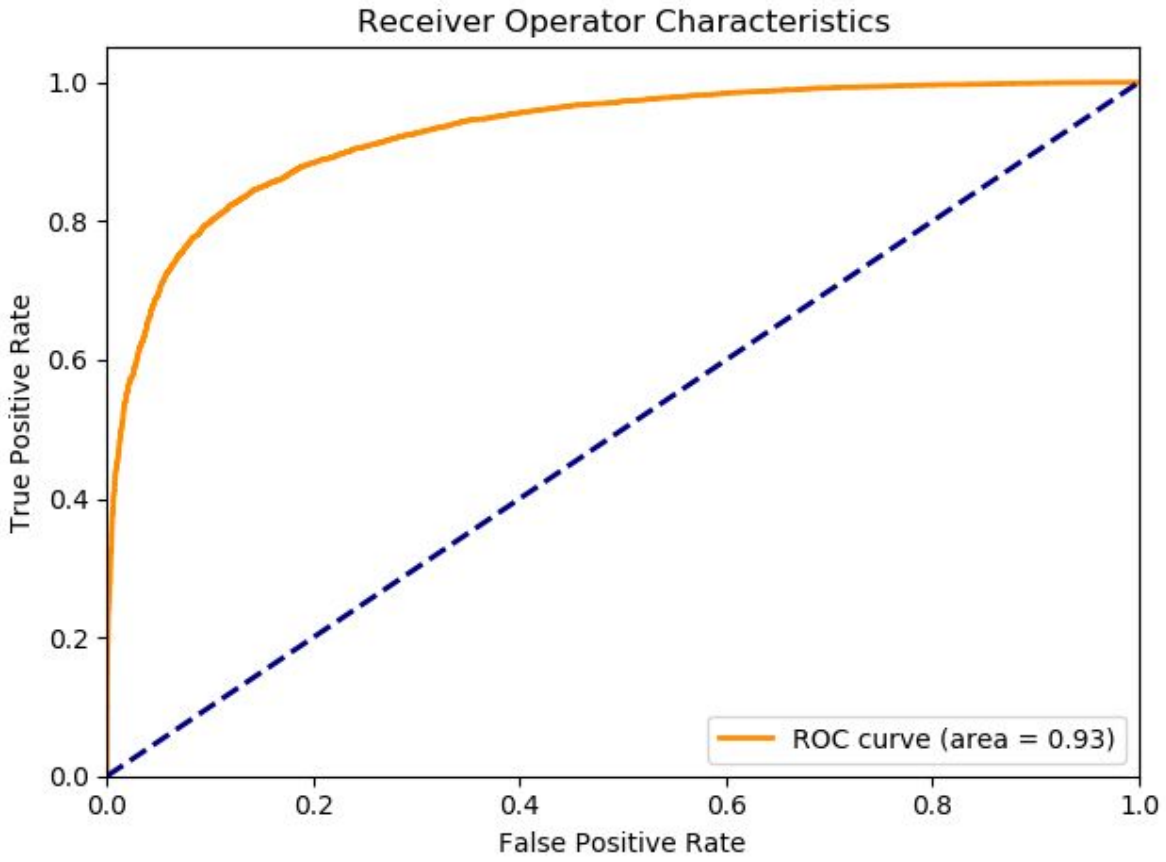

Figure 1 - ROC curve and AUC of the given convolutional neural network

Exercise 2
It is possible to construct a neural network model that is equivalent to the model above, however, only using convolutional layers (i.e. without using any fully connected or "dense" layers). Construct and train such a model.

What would be the advantage of only using convolutional layers?

Convolutional neural networks (CNN's) are more memory efficient, as they use weight sharing. This results in much fewer parameters than a simple neural network, so a computer can much more easily and faster run an analysis of CNN's. Besides that, they are able to learn relevant features from an image at different levels, like a human brain. This is called feature learning. Conventional neural networks cannot do this either. [3]

The construction of the model, consisting only of convolutional layers, can be found in the .zip file. As output layer the single output node is replaced by a convolutional layer containing 1 node and a global average pool. The ROC curve and AUC of this model can be found in figure 2.

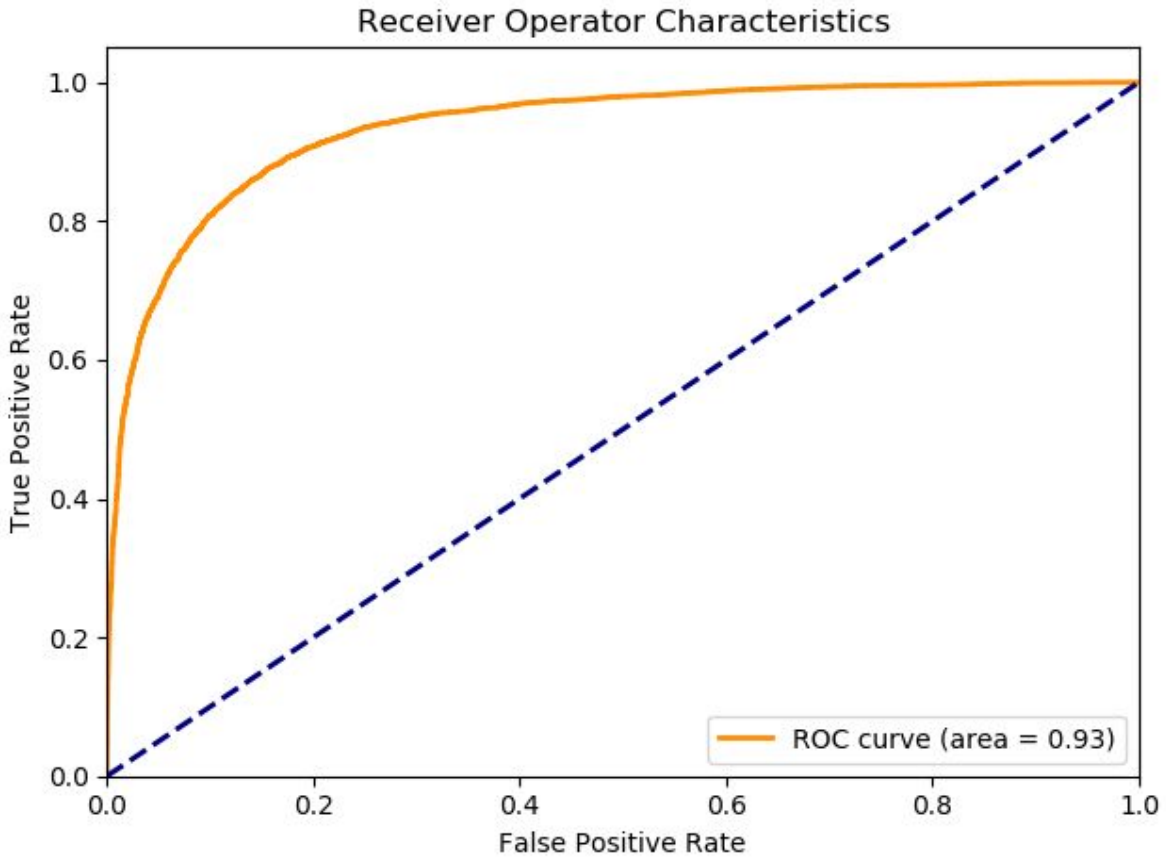

Figure 2 - ROC curve and AUC of the adapted model, only using convolutional layers

The AUC for the adaptive model is equal to the AUC from the model used in figure 1 (including a hidden dense layer). The run time on the other hand did not decrease significantly. Strangely, it even increased a bit. These runtimes are measured using the time package in python. The times were 788 seconds and 807 seconds respectively. The small difference in run times might be due to the small difference between the two models, as only two dense layers have been changed into convolutional layers. All the other layers in both models are the same, convolutional layers.

Exercise 3
Use the `kaggle_submission.py` file to prepare and submit results for evaluation to Kaggle. What is the result of the evaluation on the test set? How does it compare to the evaluation that you performed on the validation set?

The kaggle submission file (submission.csv) is included in the .zip file. The kaggle submission resulted in a AUC of 0.92. This is 0.01 lower than the AUC of the validation data found above.

References

[1] Hajian-Tilaki K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of internal medicine*, *4*(2), 627–635.

[2] Narkhede, S. (2018). Understanding AUC - ROC Curve. Website: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

[3] Advantages of AUC vs standard accuracy. Website: https://datascience.stackexchange.com/questions/806/advantages-of-auc-vs-standard-accuracy