AI in Medicine I
WiSe 2022/23

Technical University Munich
Department of Informatics

I31 - AI in Medicine and Healthcare
I32 - Computational Imaging in AI and Medicine

Theoretical exercise 3 **14. Nov. 2022**

Class Imbalance, data augmentation and transfer learning

The solutions will be discussed in the tutorial session

17. Nov 2022, 4-6 p.m. in lecture hall 5901.EG.051

For questions regarding this exercise sheet, please contact: `haifa.beji@tum.de`
For general questions, please contact: `course.aim-lab@med.tum.de`

1 Multiple choice questions

Answer the following multiple choice questions. For every question, at least one of the provided answers is correct. A point is rewarded only if all answers for a given question are correct.

1. (1 point) You want to classify images in three different age classes ($0-30$, $30-60$, $60+$ years) and you notice that your classifier tends to class most of the images in the class $60+$. What would you do to understand what the problem is and correct it?
   - ◯ Retrain the classifier as it is (maybe it failed the first time but it will succeed if you retrain it)
   - ◯ Check how many instance you have from each class to understand if you have some class imbalance
   - ◯ Plot the ROC curve
   - ◯ Adjust the learning rate
2. (1 point) Which of the following hold true for a ROC curve?
   - ◯ It is a plot of the false positive rate (x-axis) and the true positive rate (y-axis)
   - ◯ The classifier is able to perfectly distinguisch between all Positive and Negative class points correctly, when AUC $= 1$
   - ◯ The classifier is able to perfectly distinguisch between all Positive and the Negative class points correctly, when AUC $= 0$
   - ◯ There is a high chance that the classifier will be able to distinguisch the positive class values from negative class values when $0 <$ AUC < 0.5
   - ◯ It can be used both for binary and multi-class classification
3. (2 points) You want to classify brain images in three different age classes ($0-30$, $30-60$, $60+$ years) and you saw that you have a lot of images belonging in the class $60+$ but very few images that belong in the other classes. What would you do in this case?
   - ◯ Train a classifier that predicts the age using the data as is
   - ◯ Over-sampling to modify the training distributions

- ◯ Under-sampling to modify the training distributions
- ◯ Modify algorithm to take a class penalty or weight into consideration
- ◯ Shift the decision threshold, in order to reduce bias towards the majority class

4. (2 points) You want to classify brain images in three different age classes ($0-30$, $30-60$, $60+$ years) and you managed to correct for the class imbalance that you had before, but still you realise that your model diverges because you do not have enough data. How would you tackle this challenge?
   - ◯ Change the loss function
   - ◯ Apply some random geometric transformations to the input image data
   - ◯ Modify the image intensities
   - ◯ Use a generative model to synthesise new data from the training data
   - ◯ Retrain using a shallower network
   - ◯ Add batch normalisation

5. (2 points) You want to classify brain images in three different age classes ($0-30$, $30-60$, $60+$ years) and even after data augmentation you have not solved your problem. Do you have any idea what you could do next?
   - ◯ Use the pretained network you have that was trained to classify abdominal images in different age classes
   - ◯ Use a network trained to classify natural images (imagenet)
   - ◯ Transfer learning takes so much time, so do more image augmentation and retrain from the scratch
   - ◯ Re-use the pre-trained lower layers of a trained network and add new layers for training on your task

2 Open questions

1. (2 points) Give two examples of methods for handling class imbalance and explain briefly each method.

   .....

   .....

   .....

   .....

2. (5 points) Data augmentation is the process of synthesising more training samples if there are not enough training data available.

   (a) (2 points) Give one example for a simple technique and a generative technique for medical image data augmentation

   .....

   .....

   .....

   .....

   (b) (3 points) Draw a scheme to describe one generative data augmentation technique of your choice, explaining how the algorithm works.