

## Week 1 Quiz: Disease detection with computer vision

TOTAL POINTS 10

1. Which of the following is not one of the key challenges for AI diagnostic algorithms that is discussed in the lecture?

1 / 1 point

- ☐ Dataset size
- ☐ Multiple tasks
- ☐ Class imbalance
- ☒ Inflexible models

✓ **Correct**

This was not discussed as one of the key challenges, but more complex models can be used to fit data, to avoid underfitting.

2. You find that your training set has 70% negative examples and 30% positive. Which of the following techniques **will NOT** help for training this imbalanced dataset?

1 / 1 point

- ☐ Undersampling negative examples
- ☒ Oversampling negative examples
- 

Example	P(positive)
P1 Normal	0.6
P3 Normal	0.3
P5 Mass	0.4

- ☐ 0.00
- ☐ -0.4
- ☐ 2.19
- ☒ 1.27

✓ **Correct**

Given there are non-linear examples, the linear will be  $\sim \ln(1 - P(\text{negative}))$

- ☒ ~10 thousand to 100 thousand images

✓ **Correct**

Most often datasets will range from 10,000 to 100,000 labeled images. Fewer than 1000 is typically too few to train, validate and test a classifier, and very few datasets will have millions of images due to the cost of labeling.

5. Which of the following data augmentations would be best to apply?

1 / 1 point



- ☐



- ☒



- ☐ None of the above
- ☒ Biopsy

✓ **Correct**

Biopsy is definitely a valid method. Keep in mind that there are likely fewer data examples where patients have both the chest x-ray and an additional diagnostic test for the same disease.

- ☒ Confirmation by CT scan

✓ **Correct**

A CT scan can provide an objective ground truth. Keep in mind that there are likely fewer data examples where patients have both the chest x-ray and an additional diagnostic test for the same disease.

7. In what order should the training, validation, and test sets be sampled?

1 / 1 point

- ☒ Overly optimistic test performance
- ☐ None of the above
- ☐ Leaves too few images for the training set
- ☐ Leaves too few images for the test set

✓ **Correct**

Having images from the same patient is bad because it has been shown that the model may learn patient-specific features that are not generalizable to other patients.

9. Let's say you have a relatively small training set (~5 thousand images). Which training strategy makes the most sense?

1 / 1 point

- ☐ Train a model with randomly initialized weights
- ☒ Training a model with randomly initialized weights.

✓ **Correct**

Given a very large dataset, you have the option of training a new model instead of using a pre-trained model.

- ☐ Retraining the last layer of a pretrained model
- ☒ Retraining all layers of a pretrained model

✓ **Correct**

Given the large dataset, you have the option of training all layers of a pre-trained model. Using a pre-trained model may be faster than training a model from randomly initialized weights.

- ☐ Retraining the first layer of a pretrained model