# 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**

- ○ Class imbalance
- ○ Dataset size
- ◉ Inflexible models
- ○ Multiple tasks

✓ **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**

- ○ Oversampling positive examples
- ○ Reweighting examples in training loss
- ○ Undersampling negative examples
- ◉ Oversampling negative examples

✓ **Correct**
Given that the model is being trained on more negative examples, sampling even more negative samples will bias the model even more towards making a negative prediction.

**3.** What is the total loss from the normal (non-mass) examples in this example **1 / 1 point**

dataset?

Please use the natural logarithm in your calculation. When you use numpy.log, this is using the natural logarithm. Also, to get the total loss, please add up the losses from each 'normal' example.

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

- (●) 1.27
- ( ) 0.00
- ( ) 2.19
- ( ) -0.4

✓ **Correct**

Since these are negative examples, the losses will be $-log(1 - P(positive))$.

For P1, $-log(1 - 0.6) = 0.91.$

For P3 $-log(1 - 0.3) = 0.36.$

The sum is $0.91 + 0.36 = 1.27.$

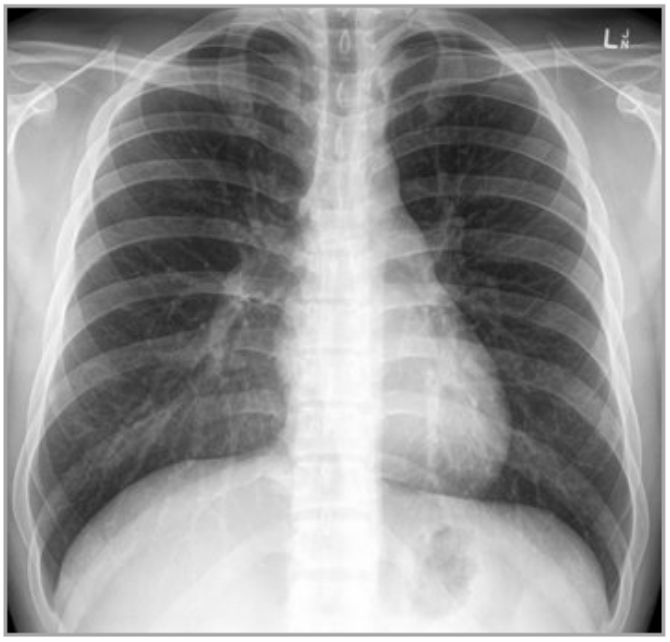
**4.** What is the typical size of medical image dataset? **1 / 1 point**

- ( ) ~1 to 1 hundred images
- ( ) ~ 1 hundred to 1 thousand images
- (●) ~10 thousand to 100 thousand images
- ( ) ~1 million or more 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

**Correct**
This rotation is most likely to help. This is a realistic transformation. Also, it does not risk changing the label.

**6.** Which of the following are valid methods for determining ground truth? Choose all that apply. **1 / 1 point**

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.

Consensus voting from a board of doctors

**Correct**
Consensus is considered less reliable than biopsy verification. However, the limited availability of biopsy data means that consensus voting may still be the best (or only viable) option.

**7.** In what order should the training, validation, and test sets be sampled? **1 / 1 point**

- ○ Validation, Test, Training
- ○ Validation, Training, Test
- ◉ Test, Validation, Training
- ○ Training, Validation, Test

**Correct**
First the test dataset should be sampled, then the validation set, then the training set. This is so that you can make sure you can adequately sample the test set, and then sample the validation set to match the distribution of labels in the test set.

**8.** Why is it bad to have the same patients in both training and test sets? **1 / 1 point**

- ◉ Overly optimistic test performance
- ○ Leaves too few images for the test set
- ○ Leaves too few images for the training set
- ○ None of these above

**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? **0 / 1 point**

- ☑ Retraining all layers of a pre-trained model

**This should not be selected**
If you retrain all the layers, a small dataset might not have enough data to generalize.

- ☐ Train a model with randomly initialized weights

- [ ] Retraining the first layer of a pre-trained model
- [x] Retraining the last layer of a pre-trained model

**Correct**
By using a pre-trained model, you can make use of its ability to recognize lower level features, and then fine tune the last few layers using your dataset.

**10.** Now let's say you have a very large dataset (~1 million images). Which training strategies will make the most sense? **0.75 / 1 point**

- [x] 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 last layer of a pretrained model
- [x] Retraining the first layer of a pretrained model

**This should not be selected**
Retraining the first layer of a pre-trained model would train the model on lower level features, but not necessarily help the model learn the higher level features of your dataset.

- [x] 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.