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 point

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

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 point

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

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

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

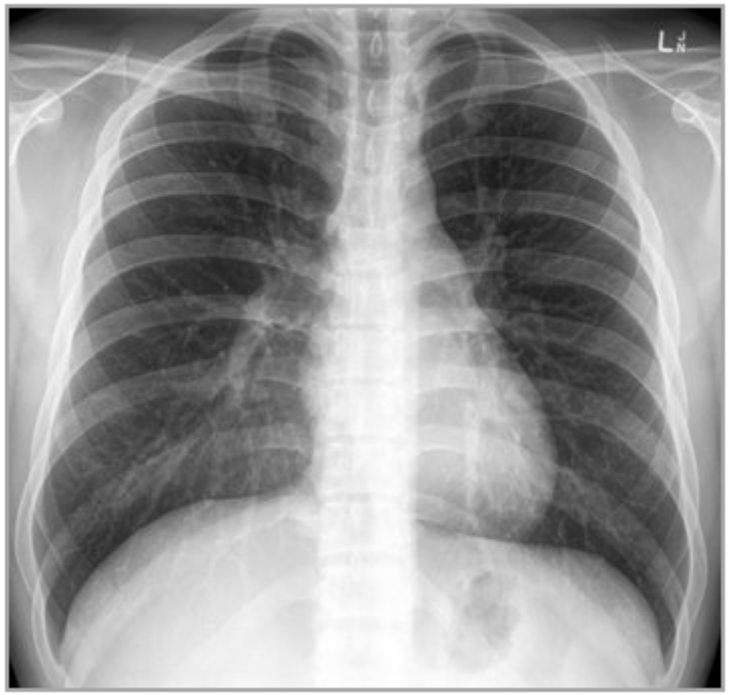
- ( ) -0.4
- ( ) 1.27
- ( ) 0.00
- (•) 2.19

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

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

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

1 point



None of the above

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

- [x] Biopsy
- [x] Confirmation by CT scan
- [x] Consensus voting from a board of doctors

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

- (●) Test, Validation, Training
- ( ) Validation, Training, Test
- ( ) Validation, Test, Training
- ( ) Training, Validation, Test

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

- (●) Overly optimistic test performance
- ( ) Leaves too few images for the test set
- ( ) Leaves too few images for the training set
- ( ) None of the above

9. Let’s say you have a relatively small training set (~5 thousand images). Which training strategy makes the most sense? 1 point

- [x] Retraining the last layer of a pre-trained model
- [ ] Retraining all layers of a pre-trained model
- [ ] Train a model with randomly initialized weights
- [ ] Retraining the first layer of a pre-trained model

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

- [ ] Training a model with randomly initialized weights.
- [x] Retraining all layers of a pretrained model
- [ ] Retraining the first layer of a pretrained model
- [ ] Retraining the last layer of a pretrained model