Congratulations! You passed!
TO PASS 70% or higher

Keep Learning

GRADE
100%

Phase 2. Project 1

LATEST SUBMISSION GRADE
100%

1. The team split the data into two partitions: the training set and the test set. It is considered best practice to have a third partition– the validation set. What added utility is there in having a validation set? **Check all that apply**. **1 / 1 point**

- [ ] The validation set can be used for updating the model directly
- [x] The validation set can be used for early stopping

Correct
The validation set should be used to roughly measure generalization performance, and therefore should be used to determine the best model.

- [x] The validation set can be used for tuning hyperparameters

Phase 2. Project 1
Graded Quiz • 30 min
Due Mar 22, 2:59 PM CST

- [x] The test set should be used for final evaluation only

Correct
The test set should not be touched during training, because we want the most realistic evaluation of the model possible.

2. The team split the data randomly, without accounting for the patient to whom each exam belongs. Why would this be a problem? *Recall: "The COVID dataset consists of 30,000 exams across 21,000 patients (some patients may be associated with multiple exams)"* **1 / 1 point**

- ( ) Patient overlap between the training and test sets may lead to problems with model bias because of the underrpresentation of certain patient demographics in the training set
- (•) Patient overlap between the training and test sets may lead to inflated model performance due to unrealistic evaluation conditions
- ( ) Patient overlap between the training and test sets may lead to problems with model convergence due to exposure to the test set
- ( ) Patient overlap between the training and test sets may lead to the leakage of PHI or other sensitive data

Correct
The features of a given patient may span across multiple exams, which may allow the model to recognize features on what should ostensibly be unseen exams and unseen patients. This may lead to inflated model performance.

3. The team downsized the images to 224 by 224 pixels. Why might this lead to worse model performance? **1 / 1 point**

Phase 2. Project 1
Graded Quiz • 30 min
Due Mar 22, 2:59 PM CST

- (•) The discriminative features in the image may be too small to identify without a higher resolution
- ( ) Memory constraints may limit the model's ability to process high-resolution images

Correct
The indicators of COVID in CXR images are likely to be small, and downsizing the samples too much will lead

Phase 2. Project 1
Graded Quiz • 30 min
Due Mar 22, 2:59 PM CST

4. Why are Convolutional Neural Networks (CNN) particularly well suited for image classification tasks? **Check all that apply**. **1 / 1 point**

- [ ] CNN architectures leverage multiple decision trees in order to make their predictions more robust
- [x] CNN architectures take advantage of feature locality through the use of filters

Correct
Convolutional filters are typically small squares of weights, therefore each pixel is evaluated alongside its neighboring pixels.

- [ ] CNN architectures can condition on previous timesteps, which it takes as input in addition to the images themselves
- [x] CNN architectures are parameter-efficient because they use the same set of weights on each region of the image

Correct
Convolutional filters are a set of weights that scan over the entire image, therefore they are used to process all regions of the image.

5. What learning phenomena is the team observing? **1 / 1 point**

- (•) Underfitting
- ( ) Generalization
- ( ) Convergence
- ( ) Overfitting