14 observations (labels):
label_names = ['No Finding', 'Enlarged Cardiomediastinum', 'Cardiomegaly', 'Lung Opacity', 'Lung Lesion', 'Edema', 'Consolidation', 'Pneumonia', 'Atelectasis', 'Pneumothorax', 'Pleural Effusion', 'Pleural Other', 'Fracture', 'Support Devices']

3-Class (0: negative, 1: positive, 2: uncertain):
https://arxiv.org/pdf/1901.07031.pdf

2-Class model (0: negative, 1: positive, here we simply merge uncertain into positive for training):
https://arxiv.org/abs/1705.02315 Wang, Xiaosong, Peng, Yifan, Lu, Le, Lu, Zhiyong, Bagheri, Mohammadhadi, and Summers, Ronald M. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. arXiv preprint arXiv:1705.02315, 2017.

Input:
224x224 image, convert to RGB, random horizontal flip, normalized based on the mean and standard deviation of training dataset of ImageNet

CNN Model:
densenet121 https://arxiv.org/abs/1608.06993
initialize parameters from the model pre-trained on ImageNet:
http://www.image-net.org/papers/imagenet_cvpr09.pdf Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. p. 248–55.

Bottleneck Features: 1x1024

3-Class Output:
dense layer: 14x3, {p_0, p_1, p_2} on each label, without Softmax(), since we use the loss function CrossEntropyLoss()

Loss Function (14-label, 3-class):
for 3 classes on each label, we use CrossEntropyLoss(), which includes Softmax() and NLLLoss(). Then we take the average over 14 labels.

Final Output: apply Softmax() on only {p_0, p_1}, then use p_1 as the output of each label.

2-Class Output
dense layer: 14x1, only {p_1} on each label, without Sigmoid(), since we use the loss function BCEWithLogitsLoss()

Loss Function (14-label, 2-class):
we use BCEWithLogitsLoss(), which includes Sigmoid() and BCELoss(). Then we take the average over 14 labels.

Final Output: apply Softmax() on {p_1}

AUC in **Test**
use 2-class {p_0, p_1}, but don't include uncertain,
we have 14 AUC/ROC for 14 observations

AUC/ROC in total:
Macro-average (equal-weight for each observation)
Micro-average (weight based on number of samples)

Optimizer
Adam: β1 = 0.9 and β2 = 0.999 as default
Learning rate: 1E-4
Decayed Factor: 10 / epoch, 2 / epoch, or None (in the images attached, I use None)

Batch Size

32 for 224x224, 16 320x320

Batch Number
6

Training Time
for 224x224: ~0.5 hour / epoch
for 320x320: ~1 hour / epoch

Some AUC Comparison:

	CheXNet	CheXNeXt	CheXpert U-Ones	CheXpert 3-Class	Ours 3-Class
Atelectasis	0.8094	0.862(0.825–0.895)	0.858 (0.806,0.910)	0.821 (0.763,0.879)	0.74
Cardiomegaly	0.9248	0.831 (0.790–0.870)	0.832 (0.773,0.890)	0.854 (0.800,0.909)	0.87
Consolidation	0.7901	0.893(0.859-0.924)	0.899 (0.854,0.944)	0.937 (0.905,0.969)	0.76
Edema	0.8878	0.924(0.886-0.955)	0.941 (0.903,0.980)	0.928 (0.887,0.968)	0.87
Effusion (Pleural Effution, in CheXpert)	0.8638	0.901(0.868-0.930)	0.934 (0.901,0.967)	0.936 (0.904,0.967)	0.90
Pleural Thickening (Pleural Other, in CheXpert)	0.8062	0.798(0.744-0.849)			0.81
Pneumonia	0.7680	0.851(0.781-0.911)			0.78
Pneumothorax	0.8887	0.944(0.915-0.969)			0.87