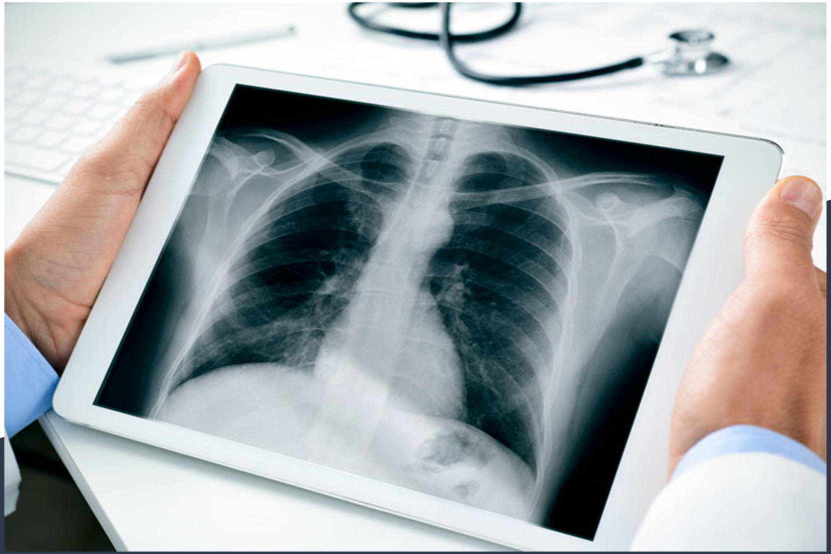
COVID-19 Radiography Database

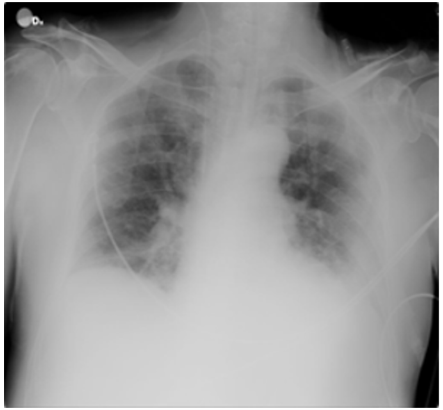


Machine Learning II
Spring 2021

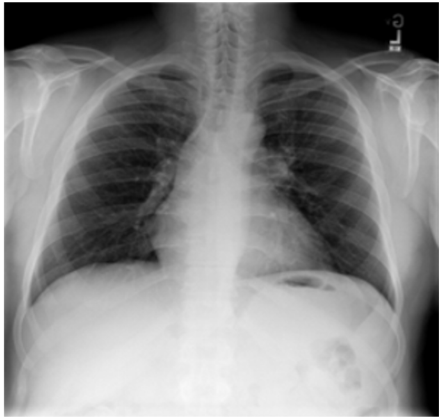
Sertan Akinci
Kristin Levine

Description of the Dataset

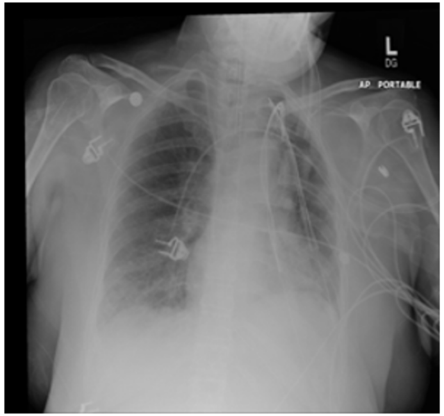
The COVID-19 Radiography Database is a large collection of four different classes of chest X-rays. 21,165 images -all .png format with resolution 299 x 299 pixels



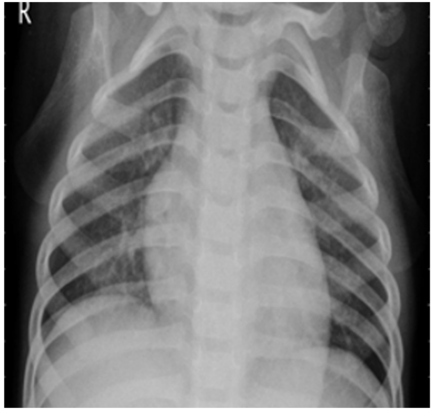
COVID-19
3616 images



Normal
10,192 images



Lung Opacity
6012 images



Viral Pneumonia
1345 images

Why this problem?

- ❖ In the midst of the pandemic, it is more important than ever to be able to quickly diagnose and treat different types of lung infections
- ❖ At a time when medical staff is greatly overworked, it would be wonderful if machine learning could help in making some of these diagnoses.

Dataset - Splitting Data

- Put aside a test set - 10% from each class
- Remaining data 80% training and 20% validation

Unbalanced classes:

COVID-19	3,254
Normal	9,173
Lung Opacity	5,410
Viral Pneumonia	1,210

Dataset – Over/Under Sample and Augment

Dataset – Balanced Data

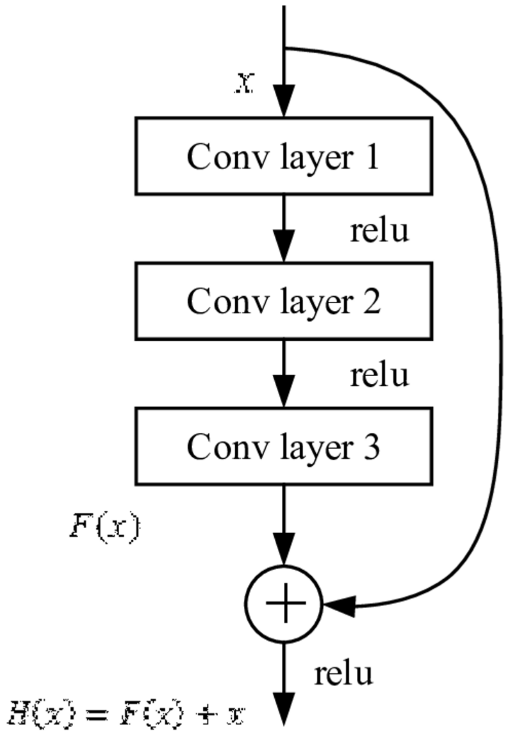
- With augmentation and undersampling:

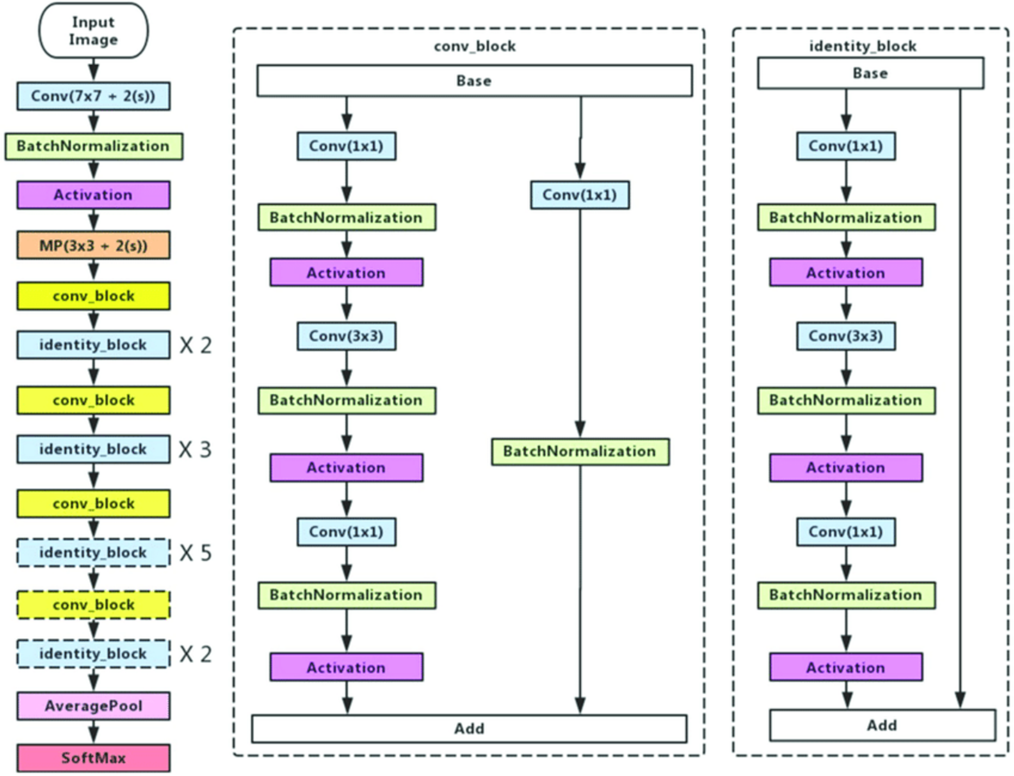
COVID-19	4,476
Normal	5,137
Lung Opacity	5,411
Viral Pneumonia	4,844

Experimental Setup

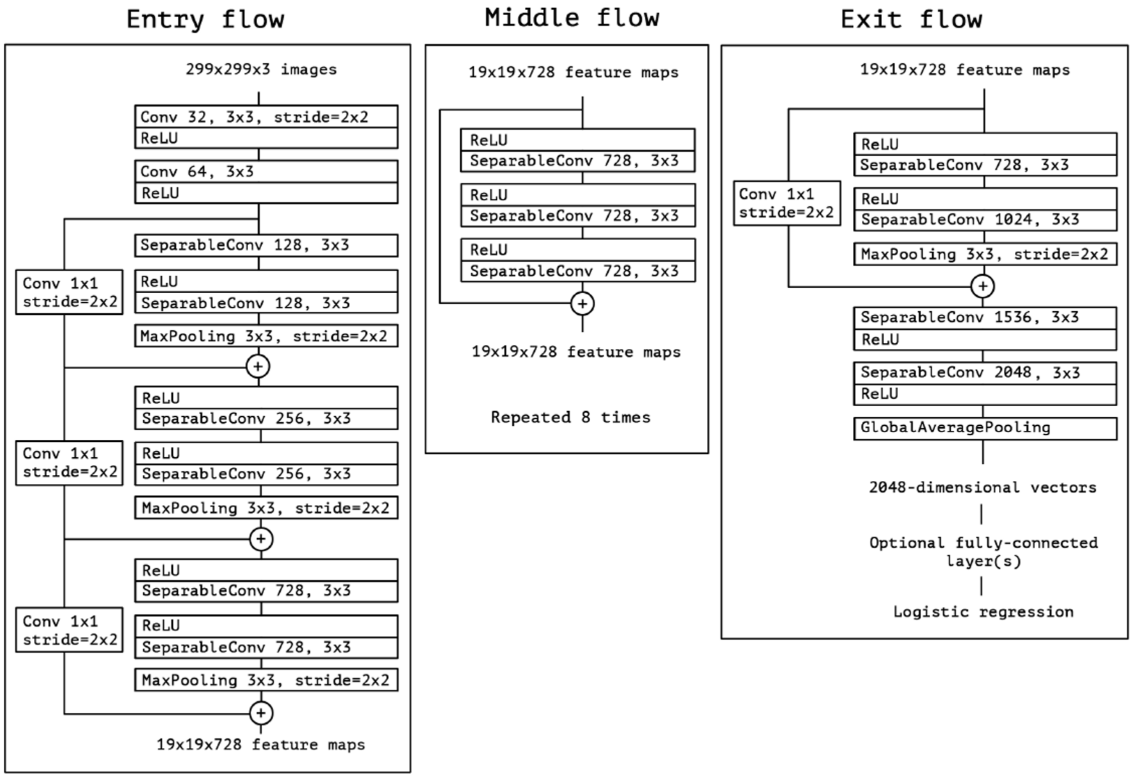
- Previous research - 2 or 3 classes
- All 4 categories; three different pre-trained models:
  - ResNet50
  - Xception
  - VGG16
- Ensemble models together using random forest algorithm

Pretrained Model – Resnet50

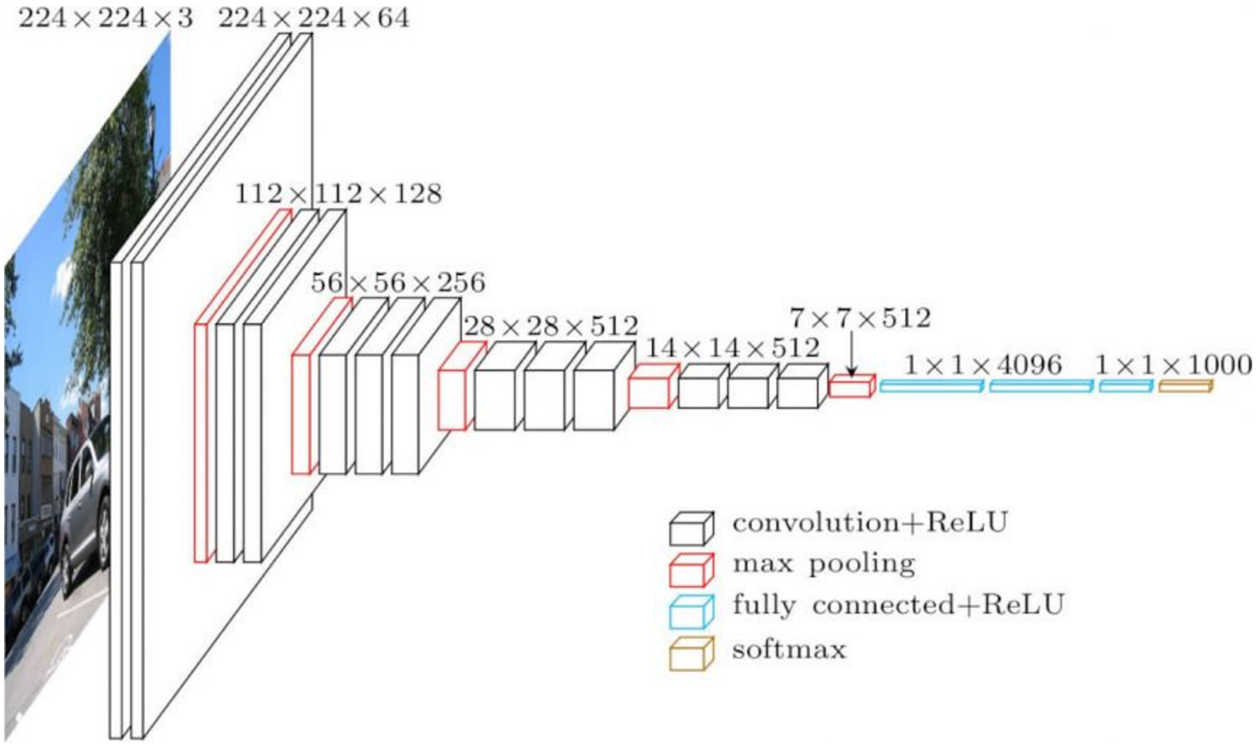




Pretrained Model – Xception



Pretrained Model – VGG16



Loading data

- Used Tensorflow’s image_dataset_from_directory
  - It labels data based on folders structure,
  - Can splits, batch, interpolate when resizing and can specify color channel

```
data_tr = tf.keras.preprocessing.image_dataset_from_directory(
     data_folder, labels='inferred',  class_names=None,
     color_mode='rgb', batch_size=batch, image_size=img_size,
     seed=random_seed,validation_split=split,
     subset='training', interpolation='bilinear')
```

Training & Parameters

- Tried freezing pretraining layers (lr = 0.001) and then unfreezing (lr =0.0001)
- Also tried training entire model along with pre-trained layers (lr = 0.0001)
- Both ways were effective !!
- Also experimented with different optimizer, Adam and SGD with momentum = 0.9

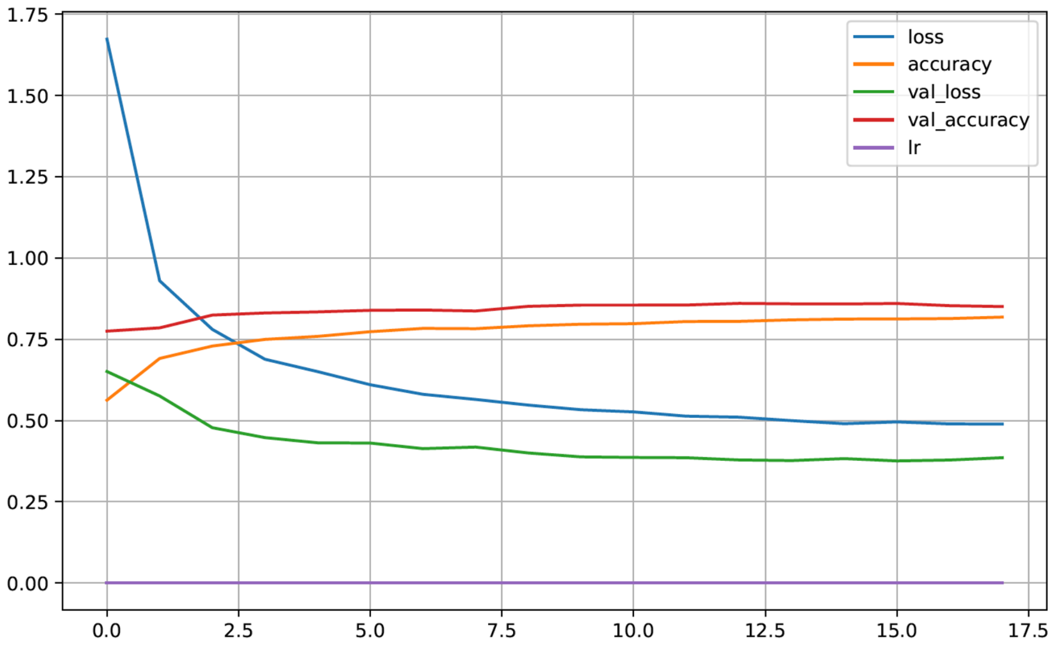
Overfitting/Early stopping

- If training accuracy higher than validation accuracy:

  Overfitting -- addressed with dropout layer

  EarlyStopping -- can use high number of epochs

Evaluating models:



- ❖ Graphed loss and accuracy vs epoch
- ❖ Also looked at precision, recall, and f1-score

Initial Results:

Training and Validation accuracy for all three models was over 90%, but for the test data it wasn't!

RESNET TEST DATA	precision	recall	f1-score	support
COVID	0.16	0.16	0.16	192
Lung_Opacity	0.31	0.30	0.30	319
Normal	0.49	0.50	0.49	541
Viral Pneumonia	0.08	0.08	0.08	71

Accuracy : 35.61887

Random Forest

Classification Report:

	precision	recall	f1-score	support
COVID	0.12	0.06	0.08	52
Lung_Opacity	0.28	0.21	0.24	98
Normal	0.50	0.72	0.59	163
Viral Pneumonia	0.00	0.00	0.00	24

Accuracy : 42.13649851632047

Loading test data wrong!

```
y_resnet = np.concatenate([y for x, y in test_data_resnet], axis=0)

And got this:
[2 2 2 ... 0 1 2]


y_xcept = np.concatenate([y for x, y in test_data_xcept], axis=0)

And got this:
[0 2 1 ... 0 1 2]
```

Corrected Testset – ResNet & Xception:

RESNET	precision	recall	f1-score	support
COVID	0.93	0.98	0.16	316
Lung_Opacity	0.93	0.89	0.30	601
Normal	0.93	0.95	0.49	1019
Viral Pneumonia	0.95	0.94	0.08	134

Accuracy : 93.8573423

XCEPTION	precision	recall	f1-score	support
COVID	0.91	0.98	0.16	316
Lung_Opacity	0.90	0.89	0.30	601
Normal	0.95	0.95	0.49	1019
Viral Pneumonia	0.95	0.94	0.08	134

Accuracy : 90.2127659

Corrected VGG16 / Random Forest

VGG16	precision	recall	f1-score	support
COVID	0.94	0.99	0.96	316
Lung_Opacity	0.92	0.92	0.92	601
Normal	0.95	0.94	0.95	1019
Viral Pneumonia	0.99	0.90	0.95	134

Accuracy : 94.2316478

Random Forest	precision	recall	f1-score	support
COVID	0.98	0.99	0.99	112
Lung_opacity	0.93	0.90	0.92	167
Normal	0.95	0.97	0.96	318
Viral Pneumonia	0.97	0.97	0.97	38

Accuracy : 95.118621102

Ensembled Model Performed the best!

Conclusions

- ❖ 90-93% accuracy when classifying four different types of chest X-ray images; Covid, Normal, Lung Opacity, and Pneumonia
- ❖ 94% accuracy using Random Forest model by linking all our models together.
- ❖ Ensembling models is definitely a technique we’ll continue to use in the future.
- ❖ Not freezing the pre-trained models performed better than expected
- ❖ Augmentation did not help as expected

Future Improvements

- ❖ if we had to do it again, we might consider separating the data into less categories like Normal, COVID, and Other
- ❖ If we were only interested in correctly identifying COVID infections (since that is a new type of infection that clinicians might have less experience with) it might make sense to lump the other types of infections into one category.
- ❖ Finally, it would also be interesting to use different types of models -- as the ones we selected performed fairly similarly.
- ❖ Could be turned into a software that health care providers can utilize anywhere in the world