

▼ IST 718 | Final Project | Playground | Group/Team 2

▼ Project Overview:

- Classifying radiology images: Normal vs Pneumonia
- [Click here to find the dataset on Kaggle](#)

The code utilizes a pre-trained computer vision model called 'ResNet18' on radiology image data. The model is then fine tuned on the data. The deep learning model is tuned with three (3) epochs with a theoretical error rate of 1.1% and is exported to a `.pkl` file for future use.

▼ Configuring Google Drive, Kaggle

- Create free Kaggle API key
 - Needed to download the dataset from Kaggle
- Upload kaggle.json to Google Drive
 - Kaggle configuration settings
- Mount Google Drive to Colab
 - Data source copy directory
- Ensure Colab has GPU settings turned on for training

```
!python -V
```

```
Python 3.9.16
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
%pwd
```

```
 '/content '
```

```
!mkdir -p
%cd '/content/drive/MyDrive/IST 718/Final Project'
```

```
mkdir: missing operand
Try 'mkdir --help' for more information.
/content/drive/MyDrive/IST 718/Final Project
```

```
%ls -a
```

```
chest-xray-pneumonia/    Final_Project_Playground.ipynb
chest-xray-pneumonia.zip  kaggle.json
```

```
import os
os.environ['KAGGLE_CONFIG_DIR'] = '/content/drive/MyDrive/IST 718/Final Project'
```

```
!kaggle datasets download -d paultimothymooney/chest-xray-pneumonia
```

```
Downloading chest-xray-pneumonia.zip to /content/drive/MyDrive/IST 718/Final
100% 2.29G/2.29G [00:31<00:00, 104MB/s]
100% 2.29G/2.29G [00:31<00:00, 78.4MB/s]
```

```
!unzip '/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumonia.zip' -d
```

Streaming output truncated to the last 5000 lines.

[illegible]

[illegible]

▼ Review Data Samples

Examples of 3 'normal' cases:

```
import os
import random
from PIL import Image

# Set the seed for the random number generator
random.seed(69)

# Set the path of the directory containing the images
dir_path = '/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumonia/ches'
```

```
# Get a list of all the image files in the directory
files = [f for f in os.listdir(dir_path) if f.endswith('.jpeg')]

# Choose 3 random files from the list
random_files = random.sample(files, 3)

# Loop through each file and print it to the screen as a thumbnail
for file in random_files:
    im = Image.open(os.path.join(dir_path, file))
    im.thumbnail((256, 256))
    im.show()
```



Examples of 3 'abnormal' cases:

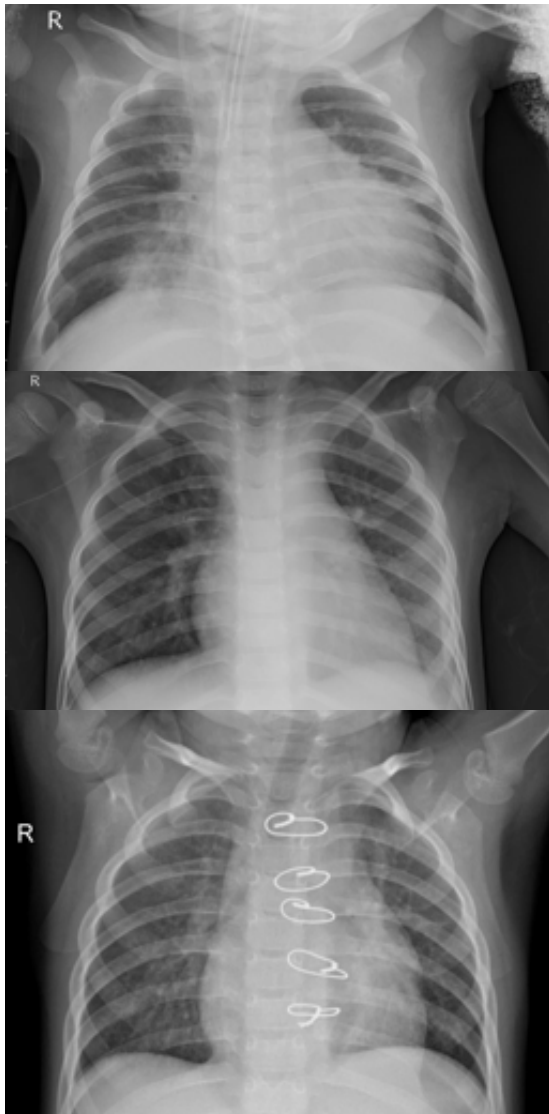
```
# Set the seed for the random number generator
random.seed(69)

# Set the path of the directory containing the images
dir_path = '/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumonia/ches

# Get a list of all the image files in the directory
files = [f for f in os.listdir(dir_path) if f.endswith('.jpeg')]

# Choose 3 random files from the list
random_files = random.sample(files, 3)

# Loop through each file and print it to the screen as a thumbnail
for file in random_files:
    im = Image.open(os.path.join(dir_path, file))
    im.thumbnail((256, 256))
    im.show()
```



▼ Exploring

- Ideas:
 - See how many images were in the dataset
 - See how big the files are (histogram of image sizes?)

▼ Train fastai Model

- Tip
 - Make sure you have your Colab GPU settings turned on...or training will take....wait for it.....forever :-)

```
from fastai.vision.all import *
```

```
path = Path('/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumonia/chest_xray/train')
path
```

```
Path('/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumonia/chest_xray/train')
```

```
dls:=DataBlock(
  ...blocks=(ImageBlock,CategoryBlock),
  ...get_items=get_image_files,
  ...splitter=RandomSplitter(valid_pct=0.2,seed=1),
  ...get_y=parent_label,
  ...item_tfms=[Resize(192,method='squish')])
dls.dataloaders(path,bs=32)
```

```
dls.show_batch(max_n=6)
```

PNEUMONIA



NORMAL



PNEUMONIA



PNEUMONIA



PNEUMONIA



NORMAL



```
learn = vision_learner(dls, resnet18, metrics=error_rate)
learn.fine_tune(3)
```

```
/usr/local/lib/python3.9/dist-packages/torchvision/models/_utils.py:208: UserWarning:
  warnings.warn(
/usr/local/lib/python3.9/dist-packages/torchvision/models/_utils.py:223: UserWarning:
  warnings.warn(msg)
Downloading: "https://download.pytorch.org/models/resnet18-f37072fd.pth" to /:
100%|██████████| 44.7M/44.7M [00:00<00:00, 165MB/s]
```

epoch	train_loss	valid_loss	error_rate	time
0	0.421595	0.138715	0.051557	02:13
epoch	train_loss	valid_loss	error_rate	time
0	0.144851	0.106104	0.026853	02:05
1	0.087914	0.045733	0.016112	02:04
2	0.038532	0.039775	0.011815	02:07

```
import os
```

```
# Save the model for future use
```

```
learn.export(f"{os.environ['KAGGLE_CONFIG_DIR']}/IST_718_Group-2_Radiology_Model_C
```

```
# Load the model
```

```
learn = load_learner(f"{os.environ['KAGGLE_CONFIG_DIR']}/IST_718_Group-2_Radiolog
learn
```

```
<fastai.learner.Learner at 0x7f175804d190>
```

▼ Test Prediction on Sample Image

```
is_normal, _, probs = learn.predict(PILImage.create('/content/drive/MyDrive/IST 718/
print(f"This is a: {is_normal}.")
print(f"Probability of normal chest x-ray without findings: {probs[0]:.4f}")
```

```
This is a: NORMAL.
```

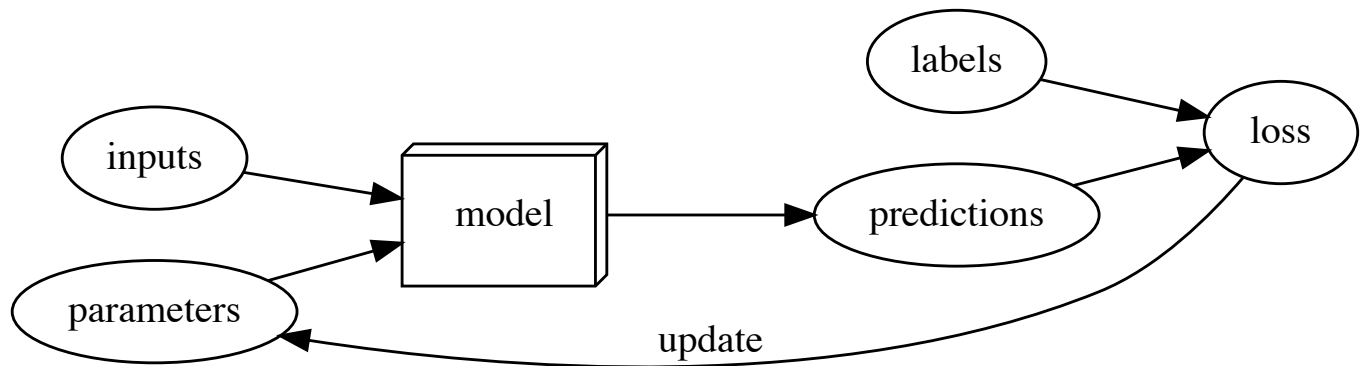
```
Probability of the patient not having pneumonia: 0.9152
```


▼ Overview of Model Architecture | Diagram

```
!pip install -Uqq graphviz
```

```
import graphviz
def gv(s):
    return graphviz.Source('digraph G{ rankdir="LR"' + s + '}; }')
```

```
gv('''ordering=in
model[shape=box3d width=1 height=0.7 label=model]
inputs->model->predictions; parameters->model; labels->loss; predictions->loss
loss->parameters[constraint=false label=update]''')
```



▼ Predict on Test Data

```
test_path = Path('/content/drive/MyDrive/IST 718/Final Project/chest-xray-pneumoni
test_files = get_image_files(test_path)
test_dl = dls.test_dl(test_files)
preds = learn.get_preds(dl=test_dl)
```

```
Exception ignored in: <function _MultiProcessingDataLoaderIter.__del__ at 0x7
Traceback (most recent call last):
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    self._shutdown_workers()
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    if w.is_alive():
  File "/usr/lib/python3.9/multiprocessing/process.py", line 160, in is_alive
    assert self._parent_pid == os.getpid(), 'can only test a child process'
AssertionError: can only test a child process
Exception ignored in: <function _MultiProcessingDataLoaderIter.__del__ at 0x7
Traceback (most recent call last):
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    self._shutdown_workers()
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    if w.is_alive():
  File "/usr/lib/python3.9/multiprocessing/process.py", line 160, in is_alive
    assert self._parent_pid == os.getpid(), 'can only test a child process'
AssertionError: can only test a child process
Exception ignored in: <function _MultiProcessingDataLoaderIter.__del__ at 0x7
Traceback (most recent call last):
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    self._shutdown_workers()
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    if w.is_alive():
  File "/usr/lib/python3.9/multiprocessing/process.py", line 160, in is_alive
    assert self._parent_pid == os.getpid(), 'can only test a child process'
AssertionError: can only test a child process
Exception ignored in: <function _MultiProcessingDataLoaderIter.__del__ at 0x7
Traceback (most recent call last):
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    self._shutdown_workers()
  File "/usr/local/lib/python3.9/dist-packages/torch/utils/data/dataloader.py
    if w.is_alive():
  File "/usr/lib/python3.9/multiprocessing/process.py", line 160, in is_alive
    assert self._parent_pid == os.getpid(), 'can only test a child process'
AssertionError: can only test a child process
```

```
pred_labels = preds[0].argmax(dim=1)
pred_labels
```

[illegible]

[illegible]

```
true_labels = [parent_label(img) for img in test_files]
correct_predictions = sum([true == pred for true, pred in zip(true_labels, pred_cl
accuracy = correct_predictions / len(test_files) * 100
```


```
for img_file, pred_class in zip(test_files, pred_class_names):
    print(f"{img_file}: {pred_class}")
```

[illegible]

[illegible]

```
print(f"Accuracy: {accuracy:.2f}%")
```

Accuracy: 83.33%

 0s completed at 3:07 PM