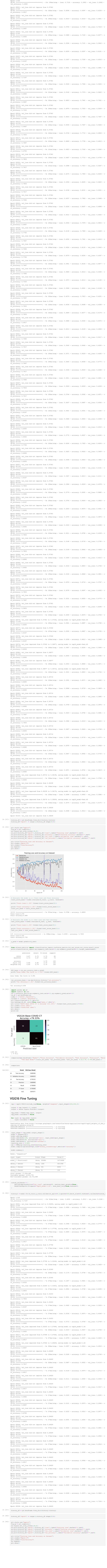
In [1]: In [2]:	<pre>from google.colab import drive drive.mount('/content/drive') Mounted at /content/drive cd /content/drive/MyDrive/ /content/drive/MyDrive</pre>
<pre>In [3]: Out[3]: In [4]:</pre>	<pre>%pwd '/content/drive/MyDrive' !git clone https://github.com/UCSD-AI4H/COVID-CT.git Cloning into 'COVID-CT' remote: Enumerating objects: 5463, done.</pre>
In [5]:	<pre>remote: Counting objects: 100% (4/4), done. remote: Compressing objects: 100% (4/4), done. remote: Total 5463 (delta 0), reused 0 (delta 0), pack-reused 5459 Receiving objects: 100% (5463/5463), 1.09 GiB 14.21 MiB/s, done. Resolving deltas: 100% (360/360), done. Checking out files: 100% (1048/1048), done. ## unzip archives q - quite, n - don't overwrite !unzip -q -n '/content/drive/My Drive/COVID-CT/Images-processed/CT_COVID.zip' !unzip -q -n '/content/drive/My Drive/COVID-CT/Images-processed/CT_NonCOVID.zip'</pre>
In [6]:	<pre>import keras import tensorflow as tf import os,cv2 import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sns from sklearn.utils import shuffle</pre>
	<pre>from sklearn.model_selection import train_test_split from keras import backend as K keras.backend.image_data_format() from keras.utils import np_utils from keras.models import Sequential from keras.layers.core import Dense, Dropout, Activation, Flatten from keras.layers.convolutional import Conv2D, MaxPooling2D from keras.layers import BatchNormalization,InputLayer from tensorflow.keras.optimizers import Adam,RMSprop,Adadelta,Adagrad,Adamax import glob</pre>
	<pre>from keras.preprocessing.image import ImageDataGenerator from keras.models import Model from keras.utils.vis_utils import plot_model from keras.callbacks import EarlyStopping,ModelCheckpoint print('Keras version is {}'.format(kerasversion)) print ('Tensorflow verson is {}'.format(tfversion)) print ('OpenCv version is {}'.format(cv2version))</pre> Keras version is 2.6.0 Tensorflow verson is 2.6.0
In [7]:	<pre>path='/content/drive/My Drive/CT_COVID/' for count, filename in enumerate(os.listdir(path)): dst ="covid-" + str(count) + ".png" src =path+ filename dst =path+ dst # rename() function will # rename all the files</pre>
	<pre>os.rename(src, dst) print("Covid Positive cases: ", count) path='/content/drive/MyDrive/CT_NonCOVID/' for count, filename in enumerate(os.listdir(path)): dst ="noncovid-" + str(count) + ".png" src =path+ filename dst =path+ dst # rename() function will</pre>
In [8]:	<pre># rename all the files os.rename(src, dst) print("Covid Negative cases: ", count) Covid Positive cases: 348 Covid Negative cases: 396 from PIL import Image</pre>
	<pre>data_path = '/content/drive/MyDrive/' #data_dir_list = os.listdir(data_path) data_dir_list = ['CT_COVID','CT_NonCOVID'] odd_size_image = 0 for dataset in data_dir_list: img_list=os.listdir(data_path+'/'+ dataset) print ('Loaded the images of dataset-'+'{}\n'.format(dataset)) for img in img_list: filename = os.path.join(data_path + '/'+ dataset + '/'+ img) image_png = Image.open(filename)</pre>
	<pre>if (image_png.size[0] or image_png.size[1])<224 : #print (img, image_png.size) odd_size_image += 1 print("number of odd sized images are {}\n".format(odd_size_image)) Loaded the images of dataset-CT_COVID number of odd sized images are 57 Loaded the images of dataset-CT_NonCOVID</pre>
In [9]: In [10]:	<pre>number of odd sized images are 77 !rm -rf /content/drive/MyDrive/COVID-CT data_path = '/content/drive/My Drive/' data_dir_list = ['CT_COVID','CT_NonCOVID'] img_rows=224</pre>
	<pre>img_cols=224 num_channel=3 num_epoch=100 no_images=0 for dataset in data_dir_list: img_list = os.listdir(data_path + '/' + dataset) no_images = no_images+len(img_list) print(no_images) # Define the number of classes labels = np.ones((no_images,),dtype='int64')</pre>
	<pre>num_classes = 2 label_index=0 img_data_list=[] img=0 for dataset in data_dir_list: img_list=os.listdir(data_path+'/'+ dataset) print ('Loaded the images of dataset-'+'{}\n'.format(dataset)) for img in img_list: input_img=cv2.imread(data_path + '/'+ dataset + '/'+ img)</pre>
	<pre>#input_img=cv2.cvtColor(input_img, cv2.Color_BGR2GRAY) input_img_resize=cv2.resize(input_img, (224,224)) input_img_resize = cv2.normalize(input_img_resize, None, alpha=0, beta=1, norm_type=cv2.NORM_MINMAX, dt img_data_list.append(input_img_resize) if dataset[0:4] == 'CT_C': labels[label_index] = 1 #print(dataset[0]) if dataset[0:4] == 'CT_N': labels[label_index] = 0 label index = label index+1</pre>
In [11]:	746 Loaded the images of dataset-CT_COVID Loaded the images of dataset-CT_NonCOVID img_data = np.array(img_data_list) print(img_data.shape)
In [12]:	<pre>plt.figure(figsize=(5, 4)) for i in range(20): plt.subplot(4, 5, i+1) plt.imshow(img_data[i], cmap='gray') plt.xticks([]) plt.yticks([]) plt.tight_layout() plt.show()</pre>
	Clipping input data to the valid range for imshow with RGB data ([01] for floats or [0255] for integers).
In [13]:	<pre>if num_channel==1: if K.image_data_format()=='th': img_data= np.expand_dims(img_data, axis=1) #print (img_data.shape) else: img_data= np.expand_dims(img_data, axis=3) #print (img_data.shape)</pre> else:
Out[13]:	<pre>if K.image_data_format() == 'th': img_data=np.rollaxis(img_data,3,1) #print (img_data.shape) labels[0:347] = 1 labels[348:745] = 0 labels[0:745] array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1</pre>
	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
	1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
	0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
In [14]:	<pre># Define the number of classes num_classes = 2 names = ['COVID','NonCOVID'] # convert class labels to on-hot encoding Y = np_utils.to_categorical(labels, num_classes) X,y = shuffle(img_data, Y, random_state=42) X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y) print("X_train_size_is {} and y_train_size_is {}".format(X_train_shape, y_train_shape)) print("Y_test_size_is {} and y_test_size_is {}".format(Y_test_shape_y_test_shape))</pre>
In [30];	<pre>print("X_train size is {} and y_train size is {}".format(X_train.shape,y_train.shape)) print("X_test size is {} and y_test size is {}".format(X_test.shape,y_test.shape)) X_train size is (596, 224, 224, 3) and y_train size is (596, 2) X_test size is (150, 224, 224, 3) and y_test size is (150, 2) Feature Extraction using VGG16 from keras.applications import vgg16 vgg = vgg16.VGG16(include_top=True, weights='imagenet',input_shape=(224,224,3))</pre>
	block1_conv2 (Conv2D) (None, 224, 224, 64) 36928 block1_pool (MaxPooling2D) (None, 112, 112, 64) 0 block2_conv1 (Conv2D) (None, 112, 112, 128) 73856 block2_conv2 (Conv2D) (None, 112, 112, 128) 147584 block2_pool (MaxPooling2D) (None, 56, 56, 128) 0
	block3_conv1 (Conv2D) (None, 56, 56, 256) 295168 block3_conv2 (Conv2D) (None, 56, 56, 256) 590080 block3_conv3 (Conv2D) (None, 56, 56, 256) 590080 block3_pool (MaxPooling2D) (None, 28, 28, 256) 0 block4_conv1 (Conv2D) (None, 28, 28, 512) 1180160 block4_conv2 (Conv2D) (None, 28, 28, 512) 2359808
	block4_conv3 (Conv2D) (None, 28, 28, 512) 2359808 block4_pool (MaxPooling2D) (None, 14, 14, 512) 0 block5_conv1 (Conv2D) (None, 14, 14, 512) 2359808 block5_conv2 (Conv2D) (None, 14, 14, 512) 2359808 block5_conv3 (Conv2D) (None, 14, 14, 512) 2359808
	block5_pool (MaxPooling2D) (None, 7, 7, 512) 0 flatten (Flatten) (None, 25088) 0 fc1 (Dense) (None, 4096) 102764544 fc2 (Dense) (None, 4096) 16781312 predictions (Dense) (None, 1000) 4097000 ================================
In [31]:	
In [32]:	<pre>layer.trainable = False model.summary() Model: "model_1" Layer (type)</pre>
	block1_conv1 (Conv2D) (None, 224, 224, 64) 1792 block1_conv2 (Conv2D) (None, 224, 224, 64) 36928 block1_pool (MaxPooling2D) (None, 112, 112, 64) 0 block2_conv1 (Conv2D) (None, 112, 112, 128) 73856 block2_conv2 (Conv2D) (None, 112, 112, 128) 147584 block2_pool (MaxPooling2D) (None, 56, 56, 128) 0
	block3_conv1 (Conv2D) (None, 56, 56, 256) 295168 block3_conv2 (Conv2D) (None, 56, 56, 256) 590080 block3_conv3 (Conv2D) (None, 56, 56, 256) 590080 block3_pool (MaxPooling2D) (None, 28, 28, 256) 0 block4_conv1 (Conv2D) (None, 28, 28, 512) 1180160
	block4_conv2 (Conv2D) (None, 28, 28, 512) 2359808 block4_conv3 (Conv2D) (None, 28, 28, 512) 2359808 block4_pool (MaxPooling2D) (None, 14, 14, 512) 0 block5_conv1 (Conv2D) (None, 14, 14, 512) 2359808 block5_conv2 (Conv2D) (None, 14, 14, 512) 2359808 block5_conv3 (Conv2D) (None, 14, 14, 512) 2359808
	block5_pool (MaxPooling2D) (None, 7, 7, 512) 0 flatten (Flatten) (None, 25088) 0 fc1 (Dense) (None, 4096) 102764544 fc2 (Dense) (None, 4096) 16781312 dense_1 (Dense) (None, 2) 8194
In [33]: In [25]:	Total params: 134,268,738 Trainable params: 8,194 Non-trainable params: 134,260,544 model.compile(optimizer='Adam',loss='binary_crossentropy',metrics=['accuracy']) plot_model(model,show_dtype=True,show_layer_names=True,show_shapes=True,rankdir='TB')
Out[25]:	input_1: InputLayer float32 input: [(None, 224, 224, 3)] output: [(None, 224, 224, 3)] block1_conv1: Conv2D float32 input: (None, 224, 224, 3) autput: (None, 224, 224, 3)
	output: (None, 224, 224, 64) block1_conv2: Conv2D float32 input: (None, 224, 224, 64) output: (None, 224, 224, 64)
	block1_pool: MaxPooling2D
	output: (None, 112, 112, 128) block2_conv2: Conv2D float32 input: (None, 112, 112, 128) output: (None, 112, 112, 128) output: (None, 112, 112, 128)
	block2_pool: MaxPooling2D float32 input: (None, 112, 112, 128) output: (None, 56, 56, 128) block3_conv1: Conv2D float32 input: (None, 56, 56, 128)
	output: (None, 56, 56, 256)
	block3_conv3: Conv2D
	block3_pool: MaxFooling2D noat32 output: (None, 28, 28, 256)
	block4_conv2: Conv2D float32 input: (None, 28, 28, 512) output: (None, 28, 28, 512) block4_conv3: Conv2D float32 input: (None, 28, 28, 512)
	output: (None, 28, 28, 512)
	block5_conv1: Conv2D
	output: (None, 14, 14, 512)
	block5_pool: MaxPooling2D float32 input: (None, 14, 14, 512) output: (None, 7, 7, 512) flatten: Flatten float32 input: (None, 7, 7, 512) output: (None, 25088)
	fc1: Dense float32 input: (None, 25088) output: (None, 4096)
	fc2: Dense float32 input: (None, 4096) output: (None, 4096) dense: Dense float32 input: (None, 4096) output: (None, 4096) output: (None, 2)
In [26]: In [35]:	<pre>custom_callbacks = [EarlyStopping(monitor="val_loss", patience=20, restore_best_weights=True), ModelCheckpoint(filepath='vgg16_model-fext.h5', save_best_only=True, verbose=1)] #del history</pre>
In [36]:	<pre>!rm vgg16_model-fext.h5 history = model.fit(X_train,y_train,validation_split=0.2,epochs=100,batch_size=10,verbose=1,callbacks=custom_cal</pre>



	0.0 - 0.0 -
n [78]:	<pre># Evaluate the model w.r.t Train Loss and Train Accuracy train_score_vgg16 = model.evaluate(X_train, y_train, verbose=1) print('Train Loss:{:.3f}'.format(train_score_vgg16[0])) print('Train accuracy:{:.3f}'.format(train_score_vgg16[1])) train_acc_fine = train_score_vgg16[1] 19/19 [====================================</pre>
n [72]: n [73]: n [74]:	from sklearn.metrics import classification_report,confusion_matrix,roc_auc_score,roc_curve,recall_score print(classification_report(np.argmax(y_test,axis=-1),np.argmax(y_pred,axis=-1),target_names=names)) precision recall f1-score support COVID 0.91 0.94 0.93 80 NonCOVID 0.93 0.90 0.91 70 accuracy 0.92 150 macro avg 0.92 0.92 0.92 150 weighted avg 0.92 0.92 0.92 150
n [75]:	print("Area Under the Curve is {:.3f}".format(AUC_vgg16)) Area Under the Curve is 0.956
	VGG16 Fine Tune COVID CT
[79]: [80]: t[80]:	Predicted label
	4 Recall 0.790000 0.920000 5 F1 Score 0.795000 0.920000 6 AUC 0.891071 0.956429