```
In [30]: import os
         import alob
         import h5pv
         import shutil
         import imgaug as aug
         import numpy as np # linear algebra
         import pandas as pd # data processing, CSV file I/O (e.g. pd.read csv)
         import seaborn as sns
         import matplotlib.pyplot as plt
         import matplotlib.image as mimg
         import imgaug.augmenters as iaa
         from os import listdir, makedirs, getcwd, remove
         from os.path import isfile, join, abspath, exists, isdir, expanduser
         from PIL import Image
         from pathlib import Path
         from skimage.io import imread
         from skimage.transform import resize
         from keras.models import Sequential, Model
         from keras.applications.vgq16 import VGG16, preprocess input
         from keras.preprocessing.image import ImageDataGenerator, load img, img to array
         from keras.models import Sequential
         from keras.layers import Conv2D, MaxPooling2D, Dense, Dropout, Input, Flatten, SeparableConv2D
         from keras.layers import GlobalMaxPooling2D
         from keras.layers.normalization import BatchNormalization
         from keras.layers.merge import Concatenate
         from keras.models import Model
         from keras.optimizers import Adam, SGD, RMSprop
         from keras.callbacks import ModelCheckpoint, Callback, EarlyStopping
         from keras.utils import to categorical
         from sklearn.model selection import train test split
         from sklearn.preprocessing import StandardScaler
         from mlxtend.plotting import plot confusion matrix
         from sklearn.metrics import confusion matrix
         import cv2
         from keras import backend as K
         color = sns.color palette()
         %matplotlib inline
         from tensorflow.python.client import device lib
         print(device lib.list local devices())
         # Input data files are available in the "../input/" directory.
```

For example, running this (by clicking run or pressing Shift+Enter) will list the files in the input direct print(os.listdir("input"))

```
[name: "/device:CPU:0"
device_type: "CPU"
memory_limit: 268435456
locality {
}
incarnation: 11681254305549916931
, name: "/device:GPU:0"
device_type: "GPU"
memory_limit: 5337382912
locality {
   bus_id: 1
   links {
    }
}
incarnation: 10187256477293018598
physical_device_desc: "device: 0, name: GeForce GTX 1060, pci bus id: 0000:01:00.0, compute capability: 6.1"
]
['xray-best-model', 'chest_xray', 'vgg16']
```

```
In [31]: import tensorflow as tf
         # Set the seed for hash based operations in python
         os.environ['PYTHONHASHSEED'] = '0'
         # Set the numpy seed
         np.random.seed(111)
         # Disable multi-threading in tensorflow ops
         session conf = tf.ConfigProto(intra op parallelism threads=1, inter op parallelism threads=1)
         # Set the random seed in tensorflow at graph level
         tf.set random seed(111)
         # Define a tensorflow session with above session configs
         sess = tf.Session(graph=tf.get default graph(), config=session conf)
         # Set the session in keras
         K.set session(sess)
         # Make the augmentation sequence deterministic
         aug.seed(111)
In [32]: # Define path to the data directory
```

```
In [32]: # Define path to the data directory
    data_dir = Path('input/chest_xray')

# Path to train directory (Fancy pathlib...no more os.path!!)
train_dir = data_dir / 'train'

# Path to validation directory
val_dir = data_dir / 'val'

# Path to test directory
test_dir = data_dir / 'test'
```

```
In [33]: # Get the path to the normal and pneumonia sub-directories
         normal cases dir = train dir / 'NORMAL'
         pneumonia cases dir = train dir / 'PNEUMONIA'
         # Get the list of all the images
         normal cases = normal cases dir.glob('*.jpeg')
         pneumonia cases = pneumonia cases dir.glob('*.jpeg')
         # An empty list. We will insert the data into this list in (img path, label) format
         train data = []
         # Go through all the normal cases. The label for these cases will be 0
         for img in normal cases:
             train data.append((img,0))
         # Go through all the pneumonia cases. The label for these cases will be 1
         for img in pneumonia cases:
             train data.append((img, 1))
         # Get a pandas dataframe from the data we have in our list
         train data = pd.DataFrame(train data, columns=['image', 'label'],index=None)
         # Shuffle the data
         train data = train data.sample(frac=1.).reset index(drop=True)
         # How the dataframe looks like?
         train data.head()
```

Out[33]:

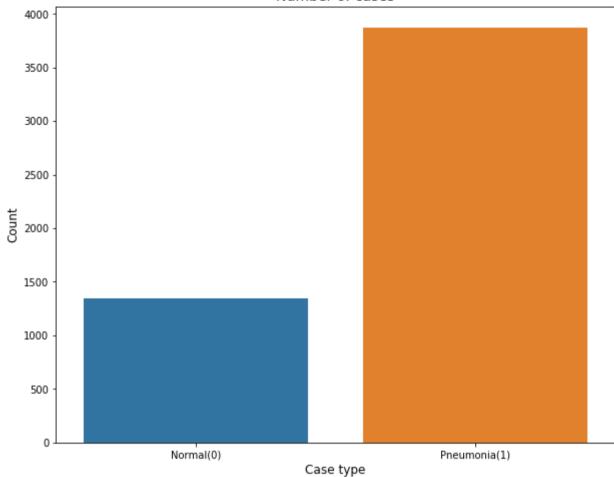
	image	label
0	input/chest_xray/train/NORMAL/NORMAL2-IM-1025	0
1	input/chest_xray/train/NORMAL/IM-0348-0001.jpeg	0
2	input/chest_xray/train/PNEUMONIA/person1468_vi	1
3	input/chest_xray/train/PNEUMONIA/person62_bact	1
4	input/chest_xrav/train/PNEUMONIA/person1503_vi	1

```
In [34]: # Get the counts for each class
    cases_count = train_data['label'].value_counts()
    print(cases_count)

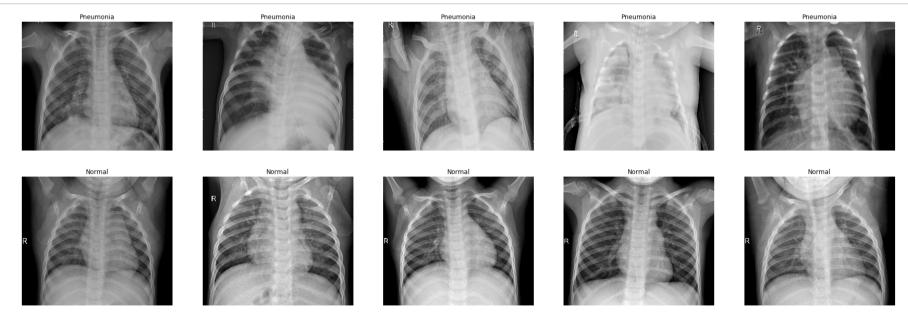
# Plot the results
    plt.figure(figsize=(10,8))
    sns.barplot(x=cases_count.index, y= cases_count.values)
    plt.title('Number of cases', fontsize=14)
    plt.xlabel('Case type', fontsize=12)
    plt.ylabel('Count', fontsize=12)
    plt.xticks(range(len(cases_count.index)), ['Normal(0)', 'Pneumonia(1)'])
    plt.show()
```

1 3875
0 1341
Name: label, dtype: int64

Number of cases



```
In [35]: # Get few samples for both the classes
         pneumonia samples = (train data[train data['label']==1]['image'].iloc[:5]).tolist()
         normal samples = (train data[train data['label']==0]['image'].iloc[:5]).tolist()
         # Concat the data in a single list and del the above two list
         samples = pneumonia samples + normal samples
         del pneumonia samples, normal samples
         # Plot the data
         f, ax = plt.subplots(2,5, figsize=(30,10))
         for i in range(10):
             img = imread(samples[i])
             ax[i//5, i%5].imshow(img, cmap='gray')
             if i<5:
                 ax[i//5, i%5].set title("Pneumonia")
             else:
                 ax[i//5, i%5].set title("Normal")
             ax[i//5, i\%5].axis('off')
             ax[i//5, i%5].set aspect('auto')
         plt.show()
```



```
In [36]: # Get the path to the sub-directories
         normal cases dir = val dir / 'NORMAL'
         pneumonia cases dir = val dir / 'PNEUMONIA'
         # Get the list of all the images
         normal cases = normal cases dir.glob('*.jpeg')
         pneumonia cases = pneumonia cases dir.glob('*.jpeg')
         # List that are going to contain validation images data and the corresponding labels
         valid data = []
         valid labels = []
         # Some images are in grayscale while majority of them contains 3 channels. So, if the image is grayscale, we
         # We will normalize the pixel values and resizing all the images to 224x224
         # Normal cases
         for img in normal cases:
             img = cv2.imread(str(img))
             img = cv2.resize(img, (224,224))
             if img.shape[2] ==1:
                 img = np.dstack([img, img, img])
             img = cv2.cvtColor(img, cv2.COLOR BGR2RGB)
             img = img.astype(np.float32)/255.
             label = to categorical(0, num classes=2)
             valid data.append(img)
             valid labels.append(label)
         # Pneumonia cases
         for img in pneumonia cases:
             img = cv2.imread(str(img))
             img = cv2.resize(img, (224,224))
             if img.shape[2] ==1:
                 img = np.dstack([img, img, img])
             img = cv2.cvtColor(img, cv2.COLOR BGR2RGB)
             img = img.astype(np.float32)/255.
             label = to categorical(1, num classes=2)
             valid data.append(img)
             valid labels.append(label)
         # Convert the list into numpy arrays
```

```
valid_data = np.array(valid_data)
valid_labels = np.array(valid_labels)

print("Total number of validation examples: ", valid_data.shape)
print("Total number of labels:", valid_labels.shape)

Total number of validation examples: (16, 224, 224, 3)
Total number of labels: (16, 2)

In [37]: # Augmentation sequence
seq = iaa.OneOf([
    iaa.Fliplr(), # horizontal flips
    iaa.Affine(rotate=20), # roatation
    iaa.Multiply((1.2, 1.5))]) #random brightness
```

```
In [38]: def data gen(data, batch size):
             # Get total number of samples in the data
             n = len(data)
             steps = n//batch size
             # Define two numpy arrays for containing batch data and labels
             batch data = np.zeros((batch size, 224, 224, 3), dtype=np.float32)
             batch labels = np.zeros((batch size,2), dtype=np.float32)
             # Get a numpy array of all the indices of the input data
             indices = np.arange(n)
             # Initialize a counter
             i =0
             while True:
                 np.random.shuffle(indices)
                 # Get the next batch
                  count = 0
                 next batch = indices[(i*batch size):(i+1)*batch size]
                 for j, idx in enumerate(next batch):
                      img name = data.iloc[idx]['image']
                     label = data.iloc[idx]['label']
                     # one hot encoding
                      encoded label = to categorical(label, num_classes=2)
                     # read the image and resize
                      img = cv2.imread(str(img name))
                      img = cv2.resize(img, (224,224))
                     # check if it's grayscale
                      if img.shape[2]==1:
                          img = np.dstack([img, img, img])
                     # cv2 reads in BGR mode by default
                      orig img = cv2.cvtColor(img, cv2.COLOR BGR2RGB)
                     # normalize the image pixels
                      orig img = img.astype(np.float32)/255.
                      batch data[count] = orig img
                      batch labels[count] = encoded label
```

```
# generating more samples of the undersampled class
    if label==0 and count < batch size-2:</pre>
        aug img1 = seq.augment image(img)
        aug img2 = seq.augment image(img)
        aug_img1 = cv2.cvtColor(aug_img1, cv2.COLOR_BGR2RGB)
        aug img2 = cv2.cvtColor(aug img2, cv2.COLOR BGR2RGB)
        aug img1 = aug img1.astype(np.float32)/255.
        aug img2 = aug img2.astype(np.float32)/255.
        batch data[count+1] = aug img1
        batch labels[count+1] = encoded label
        batch data[count+2] = aug img2
        batch labels[count+2] = encoded label
        count +=2
    else:
        count+=1
    if count==batch size-1:
        break
i+=1
yield batch data, batch labels
if i>=steps:
    i=0
```

```
In [39]: def build model():
             input img = Input(shape=(224,224,3), name='ImageInput')
             x = Conv2D(64, (3,3), activation='relu', padding='same', name='Conv1 1')(input img)
             x = Conv2D(64, (3,3), activation='relu', padding='same', name='Conv1 2')(x)
             x = MaxPooling2D((2,2), name='pool1')(x)
             x = SeparableConv2D(128, (3,3), activation='relu', padding='same', name='Conv2 1')(x)
             x = SeparableConv2D(128, (3,3), activation='relu', padding='same', name='Conv2_2')(x)
             x = MaxPooling2D((2,2), name='pool2')(x)
             x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3 1')(x)
             x = BatchNormalization(name='bn1')(x)
             x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3 2')(x)
             x = BatchNormalization(name='bn2')(x)
             x = SeparableConv2D(256, (3,3), activation='relu', padding='same', name='Conv3 3')(x)
             x = MaxPooling2D((2,2), name='pool3')(x)
             x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4 1')(x)
             x = BatchNormalization(name='bn3')(x)
             x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4 2')(x)
             x = BatchNormalization(name='bn4')(x)
             x = SeparableConv2D(512, (3,3), activation='relu', padding='same', name='Conv4 3')(x)
             x = MaxPooling2D((2,2), name='pool4')(x)
             x = Flatten(name='flatten')(x)
             x = Dense(1024, activation='relu', name='fc1')(x)
             x = Dropout(0.7, name='dropout1')(x)
             x = Dense(512, activation='relu', name='fc2')(x)
             x = Dropout(0.5, name='dropout2')(x)
             x = Dense(2, activation='softmax', name='fc3')(x)
             model = Model(inputs=input img, outputs=x)
              return model
```

In [40]: model = build_model()
model.summary()

Layer (type)	Output Shape	Param #
ImageInput (InputLayer)	(None, 224, 224, 3)	0
Conv1_1 (Conv2D)	(None, 224, 224, 64)	1792
Conv1_2 (Conv2D)	(None, 224, 224, 64)	36928
pool1 (MaxPooling2D)	(None, 112, 112, 64)	0
Conv2_1 (SeparableConv2D)	(None, 112, 112, 128)	8896
Conv2_2 (SeparableConv2D)	(None, 112, 112, 128)	17664
pool2 (MaxPooling2D)	(None, 56, 56, 128)	0
Conv3_1 (SeparableConv2D)	(None, 56, 56, 256)	34176
bn1 (BatchNormalization)	(None, 56, 56, 256)	1024
Conv3_2 (SeparableConv2D)	(None, 56, 56, 256)	68096
bn2 (BatchNormalization)	(None, 56, 56, 256)	1024
Conv3_3 (SeparableConv2D)	(None, 56, 56, 256)	68096
pool3 (MaxPooling2D)	(None, 28, 28, 256)	0
Conv4_1 (SeparableConv2D)	(None, 28, 28, 512)	133888
bn3 (BatchNormalization)	(None, 28, 28, 512)	2048
Conv4_2 (SeparableConv2D)	(None, 28, 28, 512)	267264
bn4 (BatchNormalization)	(None, 28, 28, 512)	2048
Conv4_3 (SeparableConv2D)	(None, 28, 28, 512)	267264

pool4 (MaxPooling2D)	(None, 14, 14, 512)	0
flatten (Flatten)	(None, 100352)	0
fc1 (Dense)	(None, 1024)	102761472
dropout1 (Dropout)	(None, 1024)	0
fc2 (Dense)	(None, 512)	524800
dropout2 (Dropout)	(None, 512)	0
fc3 (Dense)	(None, 2)	1026

Total params: 104,197,506 Trainable params: 104,194,434 Non-trainable params: 3,072

```
In [41]: # Open the VGG16 weight file
    f = h5py.File('input/vgg16/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5', 'r')

# Select the layers for which you want to set weight.

w,b = f['block1_conv1']['block1_conv1_W_1:0'], f['block1_conv1']['block1_conv1_b_1:0']
model.layers[1].set_weights = [w,b]

w,b = f['block1_conv2']['block1_conv2_W_1:0'], f['block1_conv2']['block1_conv2_b_1:0']
model.layers[2].set_weights = [w,b]

w,b = f['block2_conv1']['block2_conv1_W_1:0'], f['block2_conv1']['block2_conv1_b_1:0']
model.layers[4].set_weights = [w,b]

w,b = f['block2_conv2']['block2_conv2_W_1:0'], f['block2_conv2']['block2_conv2_b_1:0']
model.layers[5].set_weights = [w,b]

f.close()
model.summary()
```

Layer (type)	Output Shape	Param #
ImageInput (InputLayer)	(None, 224, 224, 3)	0
Conv1_1 (Conv2D)	(None, 224, 224, 64)	1792
Conv1_2 (Conv2D)	(None, 224, 224, 64)	36928
pool1 (MaxPooling2D)	(None, 112, 112, 64)	0
Conv2_1 (SeparableConv2D)	(None, 112, 112, 128)	8896
Conv2_2 (SeparableConv2D)	(None, 112, 112, 128)	17664
pool2 (MaxPooling2D)	(None, 56, 56, 128)	0
Conv3_1 (SeparableConv2D)	(None, 56, 56, 256)	34176
bn1 (BatchNormalization)	(None, 56, 56, 256)	1024

<pre>Conv3_2 (SeparableConv2D)</pre>	(None, 56, 56, 256)	68096
bn2 (BatchNormalization)	(None, 56, 56, 256)	1024
Conv3_3 (SeparableConv2D)	(None, 56, 56, 256)	68096
pool3 (MaxPooling2D)	(None, 28, 28, 256)	Θ
Conv4_1 (SeparableConv2D)	(None, 28, 28, 512)	133888
bn3 (BatchNormalization)	(None, 28, 28, 512)	2048
Conv4_2 (SeparableConv2D)	(None, 28, 28, 512)	267264
bn4 (BatchNormalization)	(None, 28, 28, 512)	2048
Conv4_3 (SeparableConv2D)	(None, 28, 28, 512)	267264
pool4 (MaxPooling2D)	(None, 14, 14, 512)	0
flatten (Flatten)	(None, 100352)	0
fc1 (Dense)	(None, 1024)	102761472
dropout1 (Dropout)	(None, 1024)	0
fc2 (Dense)	(None, 512)	524800
dropout2 (Dropout)	(None, 512)	0
fc3 (Dense)	(None, 2)	1026
T		

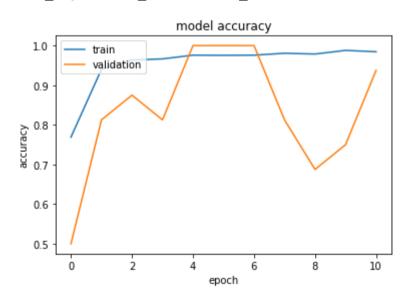
Total params: 104,197,506 Trainable params: 104,194,434 Non-trainable params: 3,072

```
In [42]: \# opt = RMSprop(lr=0.0001, decay=1e-6)
         opt = Adam(lr=0.0001, decay=1e-5)
         es = EarlyStopping(patience=5)
         chkpt = ModelCheckpoint(filepath='best model todate', save best only=True, save weights only=True)
         model.compile(loss='binary crossentropy', metrics=['accuracy'],optimizer=opt)
In [ ]:
In [43]: batch_size = 16
         nb = pochs = 20
         # Get a train data generator
         train data gen = data gen(data=train data, batch size=batch size)
         # Define the number of training steps
         nb train steps = train data.shape[0]//batch size
         print("Number of training and validation steps: {} and {}".format(nb_train_steps, len(valid_data)))
```

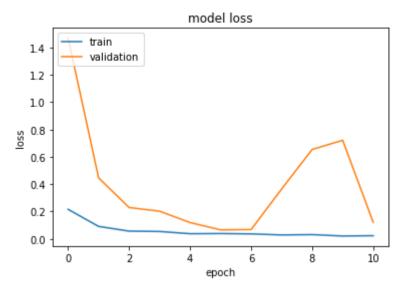
Number of training and validation steps: 326 and 16

```
In [15]: # # Fit the model
  history = model.fit generator(train data gen, epochs=nb epochs, steps per epoch=nb train steps,
   validation data=(valid data, valid labels),callbacks=[es, chkpt],
   class weight=\{0:1.0, 1:0.4\})
   Epoch 1/20
  - val acc: 0.5000
  Epoch 2/20
  - val acc: 0.8125
  Epoch 3/20
  - val acc: 0.8750
  Epoch 4/20
  - val acc: 0.8125
  Epoch 5/20
  - val acc: 1.0000
   Epoch 6/20
  - val acc: 1.0000
  Epoch 7/20
  - val acc: 1.0000
   Epoch 8/20
  - val acc: 0.8125
  Epoch 9/20
  - val acc: 0.6875
  Epoch 10/20
  - val acc: 0.7500
   Epoch 11/20
  - val acc: 0.9375
```

```
In [54]: # Load the model weights
         model.load weights("input/xray-best-model/best model.hdf5")
         print(history.history.keys())
         # "Accuracy"
         plt.plot(history.history['acc'])
         plt.plot(history.history['val acc'])
         plt.title('model accuracy')
         plt.vlabel('accuracy')
         plt.xlabel('epoch')
         plt.legend(['train', 'validation'], loc='upper left')
         plt.show()
         # "Loss"
         plt.plot(history.history['loss'])
         plt.plot(history.history['val loss'])
         plt.title('model loss')
         plt.ylabel('loss')
         plt.xlabel('epoch')
         plt.legend(['train', 'validation'], loc='upper left')
         plt.show()
```



dict keys(['val loss', 'val acc', 'loss', 'acc'])



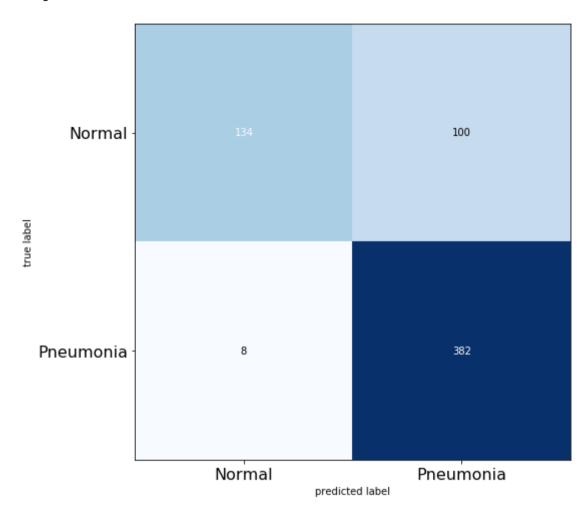
1/29/2019

```
In [55]: # Preparing test data
         normal cases dir = test dir / 'NORMAL'
         pneumonia cases dir = test dir / 'PNEUMONIA'
         normal cases = normal cases dir.glob('*.jpeg')
         pneumonia cases = pneumonia cases dir.glob('*.jpeg')
         test data = []
         test labels = []
         for img in normal cases:
              img = cv2.imread(str(img))
              img = cv2.resize(img, (224,224))
              if img.shape[2] ==1:
                  img = np.dstack([img, img, img])
              else:
                  img = cv2.cvtColor(img, cv2.COLOR BGR2RGB)
              img = img.astype(np.float32)/255.
             label = to categorical(0, num classes=2)
             test data.append(img)
             test labels.append(label)
         for img in pneumonia_cases:
              img = cv2.imread(str(img))
              img = cv2.resize(img, (224,224))
              if img.shape[2] ==1:
                  img = np.dstack([img, img, img])
              else:
                  img = cv2.cvtColor(img, cv2.COLOR BGR2RGB)
             img = img.astype(np.float32)/255.
             label = to categorical(1, num classes=2)
             test data.append(img)
             test labels.append(label)
         test data = np.array(test data)
         test labels = np.array(test labels)
         print("Total number of test examples: ", test data.shape)
         print("Total number of labels:", test labels.shape)
```

```
Total number of test examples: (624, 224, 224, 3)
         Total number of labels: (624, 2)
In [56]: # Evaluation on test dataset
         test loss, test score = model.evaluate(test_data, test_labels, batch_size=16)
         print("Loss on test set: ", test loss)
         print("Accuracy on test set: ", test score)
         624/624 [=========== ] - 4s 6ms/step
         Loss on test set: 0.9056973015925882
         Accuracy on test set: 0.8269230769230769
In [57]: # Get predictions
         preds = model.predict(test data, batch size=16)
         preds = np.argmax(preds, axis=-1)
         # Original labels
         orig test labels = np.argmax(test labels, axis=-1)
         print(orig test labels.shape)
         print(preds.shape)
         (624,)
         (624,)
```

```
In [58]: # Get the confusion matrix
cm = confusion_matrix(orig_test_labels, preds)
plt.figure()
plot_confusion_matrix(cm,figsize=(12,8), hide_ticks=True,cmap=plt.cm.Blues)
plt.xticks(range(2), ['Normal', 'Pneumonia'], fontsize=16)
plt.yticks(range(2), ['Normal', 'Pneumonia'], fontsize=16)
plt.show()
```

<Figure size 432x288 with 0 Axes>



```
In [59]: # Calculate Precision and Recall
         tn, fp, fn, tp = cm.ravel()
         precision = tp/(tp+fp)
         recall = tp/(tp+fn)
         print("Recall of the model is {:.2f}".format(recall))
         print("Precision of the model is {:.2f}".format(precision))
         Recall of the model is 0.98
         Precision of the model is 0.79
In [60]: from PIL import Image
         import numpy as np
         from skimage import transform
         def load(filename):
            np image = Image.open(filename)
            np image = np.array(np image).astype('float32')/255
            np image = transform.resize(np image, (224,224, 3))
            np_image = np.expand_dims(np_image, axis=0)
            return np image
         image = load('N1.png')
         y = model.predict(image)
         if y[0][0] <0.00001:
             print ("Normal")
         else:
             print ("Pnemonia")
```

Normal

```
In [61]: def load(filename):
    np_image = Image.open(filename)
    np_image = np.array(np_image).astype('float32')/255
    np_image = transform.resize(np_image, (224,224, 3))
    np_image = np.expand_dims(np_image, axis=0)
    return np_image

image = load('N2.png')
    y = model.predict(image)
    if y[0][0] <0.00001:
        print ("Normal")
    else:
        print ("Pnemonia")</pre>
```

Normal

```
In [62]:
    def load(filename):
        np_image = Image.open(filename)
        np_image = np.array(np_image).astype('float32')/255
        np_image = transform.resize(np_image, (224,224, 3))
        np_image = np.expand_dims(np_image, axis=0)
        return np_image

    image = load('Pl.png')
    y = model.predict(image)
    if y[0][0] <0.00001:
        print ("Normal")
    else:
        print ("Pnemonia")</pre>
```

Pnemonia

```
In [63]: def load(filename):
    np_image = Image.open(filename)
    np_image = np.array(np_image).astype('float32')/255
    np_image = transform.resize(np_image, (224,224, 3))
    np_image = np.expand_dims(np_image, axis=0)
    return np_image

image = load('P2.png')
    y = model.predict(image)
    if y[0][0] <0.00001:
        print ("Normal")
    else:
        print ("Pnemonia")</pre>
```

Pnemonia

In []: