In [1]:

```
from google.colab import drive
drive.mount('/gdrive')
%cd /gdrive/
```

```
Drive already mounted at /gdrive; to attempt to forcibly remount, call driv
e.mount("/gdrive", force_remount=True).
/gdrive
```

In [2]:

```
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from tqdm import tqdm_notebook as tqdm
from skimage import exposure
import cv2
from joblib import Parallel, delayed
!pip install pydicom
import pydicom
from pydicom import dcmread
from sklearn.utils import resample # Handle Imbalance
import pathlib
import PIL
```

```
Requirement already satisfied: pydicom in /usr/local/lib/python3.7/dist-pack
ages (2.1.2)
```

In [3]:

```
!pip install keras==2.3.1
```

```
Collecting keras==2.3.1
  Downloading https://files.pythonhosted.org/packages/ad/fd/6bfe87920d7f4fd4
75acd28500a42482b6b84479832bdc0fe9e589a60ceb/Keras-2.3.1-py2.py3-none-any.wh
l (https://files.pythonhosted.org/packages/ad/fd/6bfe87920d7f4fd475acd28500a
42482b6b84479832bdc0fe9e589a60ceb/Keras-2.3.1-py2.py3-none-any.whl) (377kB)
     |████████████████████████████████| 378kB 8.8MB/s eta 0:00:01
Requirement already satisfied: numpy>=1.9.1 in /usr/local/lib/python3.7/dist
-packages (from keras==2.3.1) (1.19.5)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.7/dist-packa
ges (from keras==2.3.1) (3.13)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.7/dist-p
ackages (from keras==2.3.1) (1.15.0)
Requirement already satisfied: keras-preprocessing>=1.0.5 in /usr/local/lib/
python3.7/dist-packages (from keras==2.3.1) (1.1.2)
Collecting keras-applications>=1.0.6
  Downloading https://files.pythonhosted.org/packages/71/e3/19762fdfc62877ae
9102edf6342d71b28fbfd9dea3d2f96a882ce099b03f/Keras_Applications-1.0.8-py3-no
ne-any.whl (https://files.pythonhosted.org/packages/71/e3/19762fdfc62877ae91
02edf6342d71b28fbfd9dea3d2f96a882ce099b03f/Keras_Applications-1.0.8-py3-none
-any.whl) (50kB)
     |████████████████████████████████| 51kB 5.8MB/s  eta 0:00:01
Requirement already satisfied: scipy>=0.14 in /usr/local/lib/python3.7/dist-
packages (from keras==2.3.1) (1.4.1)
Requirement already satisfied: h5py in /usr/local/lib/python3.7/dist-package
s (from keras==2.3.1) (2.10.0)
Installing collected packages: keras-applications, keras
  Found existing installation: Keras 2.4.3
    Uninstalling Keras-2.4.3:
      Successfully uninstalled Keras-2.4.3
Successfully installed keras-2.3.1 keras-applications-1.0.8
```

In [ ]:

```
#!pip install keras_unet
```

In [ ]:

```
#!pip install -U segmentation-models==0.2.1
```

In [4]:

```
!pip install tensorflow_io
```

```
Collecting tensorflow_io
  Downloading https://files.pythonhosted.org/packages/73/41/881ec181816767bd
91b8f2dbb319dff8eb5ff80039ed6e003c1ab8d547d7/tensorflow_io-0.17.0-cp37-cp37m
-manylinux2010_x86_64.whl (https://files.pythonhosted.org/packages/73/41/881
ec181816767bd91b8f2dbb319dff8eb5ff80039ed6e003c1ab8d547d7/tensorflow_io-0.1
7.0-cp37-cp37m-manylinux2010_x86_64.whl) (25.3MB)
     |████████████████████████████████| 25.3MB 1.4MB/s
Requirement already satisfied: tensorflow<2.5.0,>=2.4.0 in /usr/local/lib/py
thon3.7/dist-packages (from tensorflow_io) (2.4.1)
Requirement already satisfied: wheel~=0.35 in /usr/local/lib/python3.7/dist-
packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (0.36.2)
Requirement already satisfied: grpcio~=1.32.0 in /usr/local/lib/python3.7/di
st-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.32.0)
Requirement already satisfied: google-pasta~=0.2 in /usr/local/lib/python3.
7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (0.2.0)
Requirement already satisfied: gast==0.3.3 in /usr/local/lib/python3.7/dist-
packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (0.3.3)
Requirement already satisfied: protobuf>=3.9.2 in /usr/local/lib/python3.7/d
ist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.12.4)
Requirement already satisfied: keras-preprocessing~=1.1.2 in /usr/local/lib/
python3.7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.1.
2)
Requirement already satisfied: tensorflow-estimator<2.5.0,>=2.4.0 in /usr/lo
cal/lib/python3.7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_i
o) (2.4.0)
Requirement already satisfied: wrapt~=1.12.1 in /usr/local/lib/python3.7/dis
t-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.12.1)
Requirement already satisfied: numpy~=1.19.2 in /usr/local/lib/python3.7/dis
t-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.19.5)
Requirement already satisfied: termcolor~=1.1.0 in /usr/local/lib/python3.7/
dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.1.0)
Requirement already satisfied: absl-py~=0.10 in /usr/local/lib/python3.7/dis
t-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (0.12.0)
Requirement already satisfied: six~=1.15.0 in /usr/local/lib/python3.7/dist-
packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.15.0)
Requirement already satisfied: typing-extensions~=3.7.4 in /usr/local/lib/py
thon3.7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.7.4.
3)
Requirement already satisfied: h5py~=2.10.0 in /usr/local/lib/python3.7/dist
-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (2.10.0)
Requirement already satisfied: tensorboard~=2.4 in /usr/local/lib/python3.7/
dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (2.4.1)
Requirement already satisfied: astunparse~=1.6.3 in /usr/local/lib/python3.
7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.6.3)
Requirement already satisfied: flatbuffers~=1.12.0 in /usr/local/lib/python
3.7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.12)
Requirement already satisfied: opt-einsum~=3.3.0 in /usr/local/lib/python3.
7/dist-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.3.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-p
ackages (from protobuf>=3.9.2->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (54.
2.0)
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.7/d
ist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_i
o) (3.3.4)
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python
3.7/dist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorfl
ow_io) (2.23.0)
```

```
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in /usr/local/l
ib/python3.7/dist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0-
>tensorflow_io) (1.8.0)
Requirement already satisfied: werkzeug>=0.11.15 in /usr/local/lib/python3.
7/dist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow
_io) (1.0.1)
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in /usr/loca
l/lib/python3.7/dist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.
4.0->tensorflow_io) (0.4.3)
Requirement already satisfied: google-auth<2,>=1.6.3 in /usr/local/lib/pytho
n3.7/dist-packages (from tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorf
low_io) (1.28.0)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in
/usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8->tensorboard~=
2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.8.1)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.
7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.
0,>=2.4.0->tensorflow_io) (2020.12.5)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.
7/dist-packages (from requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.
0,>=2.4.0->tensorflow_io) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /u
sr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0->tensorboard~
=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.24.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist
-packages (from requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.0,>=2.
4.0->tensorflow_io) (2.10)
Requirement already satisfied: requests-oauthlib>=0.7.0 in /usr/local/lib/py
thon3.7/dist-packages (from google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=
2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (1.3.0)
Requirement already satisfied: rsa<5,>=3.1.4; python_version >= "3.6" in /us
r/local/lib/python3.7/dist-packages (from google-auth<2,>=1.6.3->tensorboard
~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (4.7.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/pytho
n3.7/dist-packages (from google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow
<2.5.0,>=2.4.0->tensorflow_io) (0.2.8)
Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/pyth
on3.7/dist-packages (from google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflo
w<2.5.0,>=2.4.0->tensorflow_io) (4.2.1)
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-pa
ckages (from importlib-metadata; python_version < "3.8"->markdown>=2.6.8->te
nsorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.4.1)
Requirement already satisfied: oauthlib>=3.0.0 in /usr/local/lib/python3.7/d
ist-packages (from requests-oauthlib>=0.7.0->google-auth-oauthlib<0.5,>=0.4.
1->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (3.1.0)
Requirement already satisfied: pyasn1>=0.1.3 in /usr/local/lib/python3.7/dis
t-packages (from rsa<5,>=3.1.4; python_version >= "3.6"->google-auth<2,>=1.
6.3->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow_io) (0.4.8)
Installing collected packages: tensorflow-io
Successfully installed tensorflow-io-0.17.0
```

In [ ]:

```
!pip list
```

In [3]:

```
import tensorflow as tf
import keras
import os
import imgaug.augmenters as iaa
from PIL import Image
# import albumentations as A
os.environ['TF_FORCE_GPU_ALLOW_GROWTH'] = 'true'
#from tensorflow.keras import layers,Model
from tensorflow.keras.layers import Dense,Input,Conv2D,MaxPool2D,Activation,Dropout,Flatten
from keras.callbacks import ModelCheckpoint, EarlyStopping, LearningRateScheduler, ReduceLR
from keras import backend as K
from tensorflow.keras.models import Model
from tensorflow.keras.losses import binary_crossentropy
import random as rn
import tensorflow_io as tfio

# we are importing the pretrained unet from the segmentation models
# https://github.com/qubvel/segmentation_models
#import segmentation_models as sm
#from segmentation_models import Unet
# sm.set_framework('tf.keras')
keras.backend.set_image_data_format('channels_last')
#from segmentation_models.metrics import iou_score
```

```
Using TensorFlow backend.
```

In [4]:

```
from glob import glob
import pathlib
import pydicom as dicom
```

In [5]:

```
os.chdir('/gdrive/MyDrive/Image_Segementation_CS2/')
```

In [6]:

```
dataset = pd.read_csv('siim/train-rle.csv')
dataset.head()
```

Out[6]:

	ImageId	EncodedPixels
0	1.2.276.0.7230010.3.1.4.8323329.6904.151787520...	-1
1	1.2.276.0.7230010.3.1.4.8323329.13666.15178752...	557374 2 1015 8 1009 14 1002 20 997 26 990 32 ...
2	1.2.276.0.7230010.3.1.4.8323329.11028.15178752...	-1
3	1.2.276.0.7230010.3.1.4.8323329.10366.15178752...	514175 10 1008 29 994 30 993 32 991 33 990 34 ...
4	1.2.276.0.7230010.3.1.4.8323329.10016.15178752...	592184 33 976 58 956 73 941 88 926 102 917 109...

In [ ]:

```
data_dir = pathlib.Path("siim")
train_read = sorted(data_dir.glob('dicom-images-train/**/*.dcm'))
test_read  = sorted(data_dir.glob('dicom-images-test/**/*.dcm'))
```

In [ ]:

```
missing_images = 0
train_df =[]
remove =[]

for j,i in enumerate(tqdm(train_read)):


  sample = dicom.dcmread(i) # reading each input
  train ={}
  train['UID'] = sample.SOPInstanceUID

  try: # try and except , to avoid throwing and error in case of missing files

    encoded_pixel = dataset[dataset['ImageId'] == train['UID']].values [0][1] # Checking ea
    train['Encoded_pixel'] = encoded_pixel
  except:
    missing_images+=1
    remove.append('siim/dicom-images-train/'+sample.StudyInstanceUID+'/'+sample.SeriesInsta
    # if the image details are not ppresent in CSV means, file is missing

  train['Path'] = 'siim/dicom-images-train/'+sample.StudyInstanceUID+'/'+sample.SeriesInsta

  train_df.append(train)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:5: TqdmDeprecat
ionWarning: This function will be removed in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
  """

HBox(children=(FloatProgress(value=0.0, max=12089.0), HTML(value='')))
```

In [ ]:

```
df_main = pd.DataFrame(train_df,columns=['UID','Encoded_pixel','Path'])
df_main.head()
```

Out[17]:

	UID	Encoded_pixel	Path
0	1.2.276.0.7230010.3.1.4.8323329.1000.151787516...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
1	1.2.276.0.7230010.3.1.4.8323329.10000.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
2	1.2.276.0.7230010.3.1.4.8323329.10001.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
3	1.2.276.0.7230010.3.1.4.8323329.10002.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
4	1.2.276.0.7230010.3.1.4.8323329.10003.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....

In [ ]:

```
df_main = df_main.loc[~df_main['Path'].isin(remove)] #remove the row which dont have image.
df_main.head()
```

Out[18]:

	UID	Encoded_pixel	Path
0	1.2.276.0.7230010.3.1.4.8323329.1000.151787516...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
1	1.2.276.0.7230010.3.1.4.8323329.10000.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
2	1.2.276.0.7230010.3.1.4.8323329.10001.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
3	1.2.276.0.7230010.3.1.4.8323329.10002.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
4	1.2.276.0.7230010.3.1.4.8323329.10003.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....

In [ ]:

```
df_downsampled = df_main[df_main['Encoded_pixel']!= '-1']
df_downsampled.drop('UID',axis=1,inplace=True)
df_downsampled.head()
```

```
/usr/local/lib/python3.7/dist-packages/pandas/core/frame.py:4174: SettingWit
hCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/
stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://pand
as.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-v
ersus-a-copy)
  errors=errors,
```

Out[53]:

	Encoded_pixel	Path
6	209126 1 1019 6 1015 10 1012 13 1010 14 1008 1...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
13	49820 3 1017 11 1012 13 1009 16 1007 18 1006 1...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
14	261328 6 1015 11 1011 15 1007 18 1004 21 1002 ...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
18	592184 33 976 58 956 73 941 88 926 102 917 109...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
28	530522 1 1022 3 1019 6 1017 7 1016 9 1014 10 1...	siim/dicom-images-train/1.2.276.0.7230010.3.1....

In [ ]:

```
df_main.to_csv('Main_CS2_SIIM_All.csv',index=False)
```

In [ ]:

```
df_downsampled.to_csv('Main_CS2_SIIM.csv',index=False)
```

In [6]:

```
# Read a saved files
df_main = pd.read_csv('Main_CS2_SIIM_All.csv')
df_downsampled = pd.read_csv('Main_CS2_SIIM.csv') # Only negative sampled.
```

In [7]:

```
df_main.head(3)
```

Out[7]:

	UID	Encoded_pixel	Path
0	1.2.276.0.7230010.3.1.4.8323329.1000.151787516...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
1	1.2.276.0.7230010.3.1.4.8323329.10000.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....
2	1.2.276.0.7230010.3.1.4.8323329.10001.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....

In [12]:

```
df_downsampled.head(3)
```

Out[12]:

	Encoded_pixel	Path
0	209126 1 1019 6 1015 10 1012 13 1010 14 1008 1...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
1	49820 3 1017 11 1012 13 1009 16 1007 18 1006 1...	siim/dicom-images-train/1.2.276.0.7230010.3.1....
2	261328 6 1015 11 1011 15 1007 18 1004 21 1002 ...	siim/dicom-images-train/1.2.276.0.7230010.3.1....

In [8]:

```
# create classfication problem
label =[]
for i in tqdm(df_main['Encoded_pixel']):
  if i =='-1':
    label.append(0)
  else:
    label.append(1)

df_main['Label'] = label
df_main.head(3)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: TqdmDeprecat
ionWarning: This function will be removed in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
  This is separate from the ipykernel package so we can avoid doing imports
until

HBox(children=(FloatProgress(value=0.0, max=12047.0), HTML(value='')))
```

Out[8]:

	UID	Encoded_pixel	Path	L
0	1.2.276.0.7230010.3.1.4.8323329.1000.151787516...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....	
1	1.2.276.0.7230010.3.1.4.8323329.10000.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....	
2	1.2.276.0.7230010.3.1.4.8323329.10001.15178752...	-1	siim/dicom-images-train/1.2.276.0.7230010.3.1....	

In [9]:

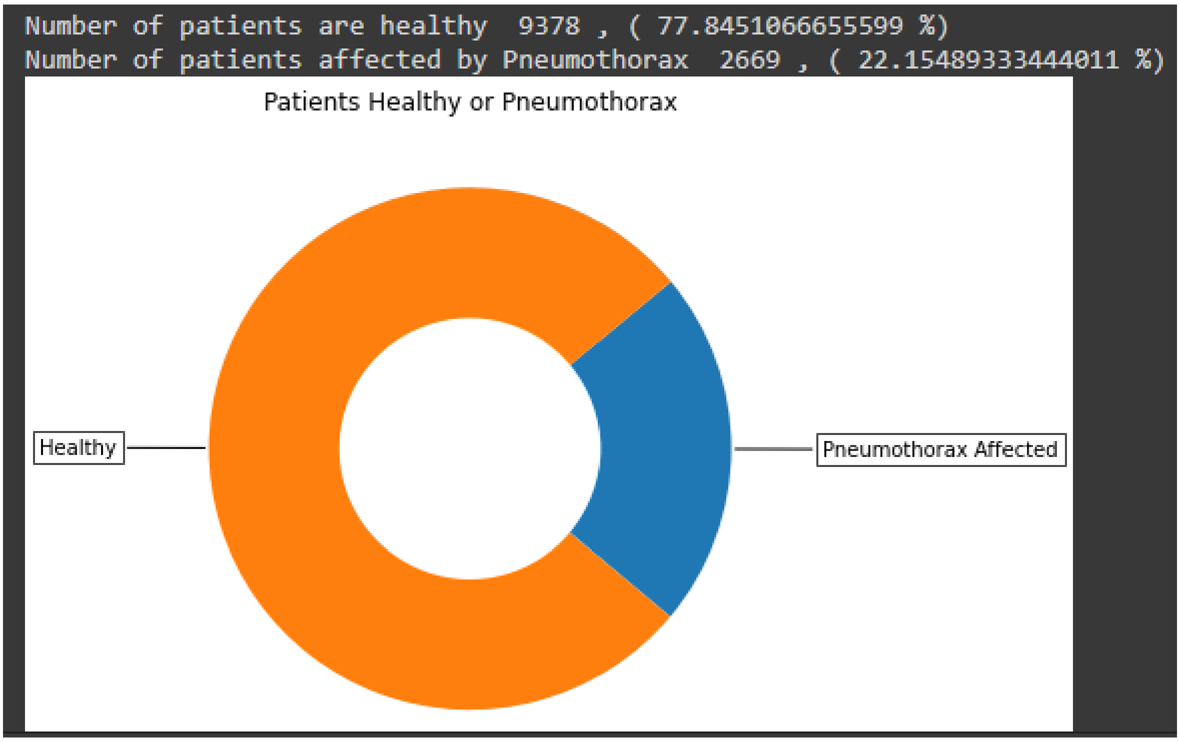
```
df_downsampled = df_main[['Path','Label']]
df_downsampled.head(3)
```

Out[9]:

	Path	Label
0	siim/dicom-images-train/1.2.276.0.7230010.3.1....	0
1	siim/dicom-images-train/1.2.276.0.7230010.3.1....	0
2	siim/dicom-images-train/1.2.276.0.7230010.3.1....	0

MODEL

Over Distribution of person affected by pneumothorax



```
Number of patients are healthy  9378 , ( 77.8451066655599 %)
Number of patients affected by Pneumothorax  2669 , ( 22.15489333444011 %)
```

In [29]:

```
file_path   = df_downsampled['Path'].values
labels      = df_downsampled['Label'].values
```

Why use Tf.Dataset pipeline?

The tf. data API enables you to build complex input pipelines from simple, reusable pieces. For example, the pipeline for an image model might aggregate data from files in a distributed file system, apply random perturbations to each image, and merge randomly selected images into a batch for training.

The Dataset API allows you to build an asynchronous, highly optimized data pipeline to prevent your GPU from data starvation. It loads data from the disk (images or text), applies optimized transformations, creates batches, and sends it to the GPU. Former data pipelines made the GPU wait for the CPU to load the data, leading to performance issues.

In [31]:

```
train_ds = tf.data.Dataset.from_tensor_slices((file_path,label))
train_ds = train_ds.shuffle(len(df_downsampled),seed=42)
train_ds
#test_ds = tf.data.Dataset.from_tensor_slices((test_path,test_mask))
```

Out[31]:

```
<ShuffleDataset shapes: ((), ()), types: (tf.string, tf.int32)>
```

In [33]:

```
def decode_img(img):
  # convert the compressed string to a 3D uint8 tensor
  #image_bytes = tf.io.read_file(img)
  image = tfio.image.decode_dicom_image(img, dtype=tf.uint8,color_dim=True,scale='preserve'

  image = tf.image.convert_image_dtype(image, tf.float32)#converting the image to tf.float3
  image=tf.squeeze(image,[0]) #squeezing the image because the file is of the shape (1,1024
  b = tf.constant([1,1,3], tf.int32)
  image=tf.tile(image,b)
  image=tf.image.resize(image,size=[256,256]) #the image is of the shape (1024,1024,1) to m
  # resize the image to the desired size
  return image

def process_path(file_path,label):
  img = tf.io.read_file(file_path) #reading the image from the file path
  img = decode_img(img) #passing the image to the function
  return img,label

AUTOTUNE = tf.data.experimental.AUTOTUNE
train_ds = train_ds.map(process_path,num_parallel_calls=AUTOTUNE)
train_ds
```

Out[33]:

```
<ParallelMapDataset shapes: ((256, 256, None), ()), types: (tf.float32, tf.i
nt32)>
```

In [34]:

```
val_size = int(len(df_downsampled)*0.2) #splitting to 80-20 data
train_ds = train_ds.skip(val_size)
test_ds = train_ds.take(val_size)
```

In [49]:

```
train_dataset_new  = train_ds.batch(64)
test_dataset_new  =  test_ds.batch(64,drop_remainder=True)
```

In [50]:

```
train_dataset_new
```

Out[50]:

```
<BatchDataset shapes: ((None, 256, 256, None), (None,)), types: (tf.float32,
tf.int32)>
```

In [51]:

```
test_dataset_new
```

Out[51]:

```
<BatchDataset shapes: ((64, 256, 256, None), (64,)), types: (tf.float32, tf.
int32)>
```

In [52]:

```
train_ds
```

Out[52]:

```
<SkipDataset shapes: ((256, 256, None), ()), types: (tf.float32, tf.int32)>
```

In [53]:

```
test_ds
```

Out[53]:

```
<TakeDataset shapes: ((256, 256, None), ()), types: (tf.float32, tf.int32)>
```

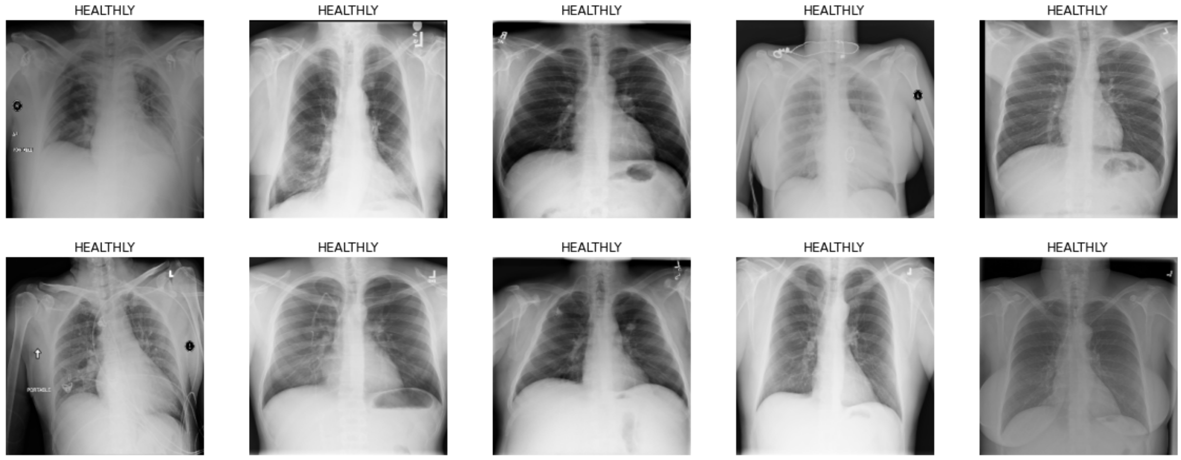
In [ ]:

```

```

In [41]:

```
plt.figure(figsize=(20,20)) #plotting images from the train as a sanity check of whether th
count=0
for i,j in tqdm(train_ds.take(10)):
  ax = plt.subplot(5,5,count+1)
  count=count+1
  if j==0:
    plt.title("HEALTHLY")
  else:
    plt.title("PNEUMOTHORAX")
  plt.imshow(i)
  plt.axis("off")
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: TqdmDeprecat
ionWarning: This function will be removed in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
  This is separate from the ipykernel package so we can avoid doing imports
until

HBox(children=(FloatProgress(value=0.0, max=10.0), HTML(value='')))
```



In [42]:

```
from tensorflow.keras import applications
```

In [43]:

```
# VGG16 as Backbone
model_1 =tf.keras.applications.vgg16.VGG16(weights = "imagenet", include_top=False, input_s
for i in model_1.layers:
  i.trainable=False
model=model_1.output
model=Conv2D(32, (3, 3))(model)
model=(Activation('relu'))(model)
model=(MaxPool2D(pool_size=(2, 2)))(model)
model=Flatten()(model)
model = Dense(256, activation="relu")(model)
model = Dense(128, activation="relu")(model)
output_layer = Dense(1, activation="sigmoid")(model)
model1 = Model(model_1.input,output_layer)
```

In [44]:

```
model1.summary()
```

```
Model: "model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 256, 256, 3)]     0
_________________________________________________________________
block1_conv1 (Conv2D)        (None, 256, 256, 64)      1792
_________________________________________________________________
block1_conv2 (Conv2D)        (None, 256, 256, 64)      36928
_________________________________________________________________
block1_pool (MaxPooling2D)   (None, 128, 128, 64)      0
_________________________________________________________________
block2_conv1 (Conv2D)        (None, 128, 128, 128)     73856
_________________________________________________________________
block2_conv2 (Conv2D)        (None, 128, 128, 128)     147584
_________________________________________________________________
block2_pool (MaxPooling2D)   (None, 64, 64, 128)       0
_________________________________________________________________
block3_conv1 (Conv2D)        (None, 64, 64, 256)       295168
_________________________________________________________________
block3_conv2 (Conv2D)        (None, 64, 64, 256)       590080
_________________________________________________________________
block3_conv3 (Conv2D)        (None, 64, 64, 256)       590080
_________________________________________________________________
block3_pool (MaxPooling2D)   (None, 32, 32, 256)       0
_________________________________________________________________
block4_conv1 (Conv2D)        (None, 32, 32, 512)       1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, 32, 32, 512)       2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 32, 32, 512)       2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 16, 16, 512)       0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 16, 16, 512)       2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 8, 8, 512)         0
_________________________________________________________________
conv2d (Conv2D)              (None, 6, 6, 32)          147488
_________________________________________________________________
activation (Activation)      (None, 6, 6, 32)          0
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 3, 3, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 288)               0
_________________________________________________________________
dense (Dense)                (None, 256)               73984
_________________________________________________________________
dense_1 (Dense)              (None, 128)               32896
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 129
=================================================================
Total params: 14,969,185
Trainable params: 254,497
Non-trainable params: 14,714,688
_________________________________________________________________
```

In [45]:

```
%load_ext tensorboard
import datetime
folder_name = datetime.datetime.now().strftime("%Y%m%d-%H%M%S")

# Create log folder - TensorBoard
log_dir="/gdrive/My Drive/Image_segmentation/segmentation/logs/fit/" + folder_name
tensorboard_callback =TensorBoard(log_dir=log_dir,histogram_freq=1, write_graph=True)

print('Folder_name', folder_name)

early_stop = keras.callbacks.EarlyStopping(
    monitor='val_recall', min_delta=0, patience=30, verbose=0, mode='auto',
    baseline=None, restore_best_weights=False
)

#Saving the best model
filepath="new_model_save_test/best_models_classification.h5"
checkpoint = keras.callbacks.ModelCheckpoint(filepath,
                                  monitor='val_recall',
                                  mode='max',
                                  verbose=1,
                                  save_best_only=True)
```

```
Folder_name 20210412-055354
```

In [57]:

```
opti = tf.keras.optimizers.Adam(lr=0.0001)
model1.compile(loss = "binary_crossentropy", optimizer =opti, metrics=["accuracy", tf.keras
model1.fit(train_dataset_new,epochs=50,verbose=1,validation_data=test_dataset_new,batch_siz
```

```
151/151 [==============================] - 346s 2s/step - loss: 0.3474 - a
ccuracy: 0.8433 - precision: 0.7226 - recall: 0.5174 - val_loss: 0.3381 -
val_accuracy: 0.8661 - val_precision: 0.7084 - val_recall: 0.7200

Epoch 00006: val_recall improved from 0.40797 to 0.72000, saving model to
new_model_save_test/best_models_classification.h5
Epoch 7/50
151/151 [==============================] - 346s 2s/step - loss: 0.3247 - a
ccuracy: 0.8583 - precision: 0.7518 - recall: 0.5383 - val_loss: 0.3202 -
val_accuracy: 0.8585 - val_precision: 0.8233 - val_recall: 0.4833

Epoch 00007: val_recall did not improve from 0.72000
Epoch 8/50
151/151 [==============================] - 346s 2s/step - loss: 0.2985 - a
ccuracy: 0.8711 - precision: 0.7768 - recall: 0.5745 - val_loss: 0.2860 -
val_accuracy: 0.8872 - val_precision: 0.7815 - val_recall: 0.6782

Epoch 00008: val_recall did not improve from 0.72000
Epoch 9/50
151/151 [==============================] - 345s 2s/step - loss: 0.2941 - a
```

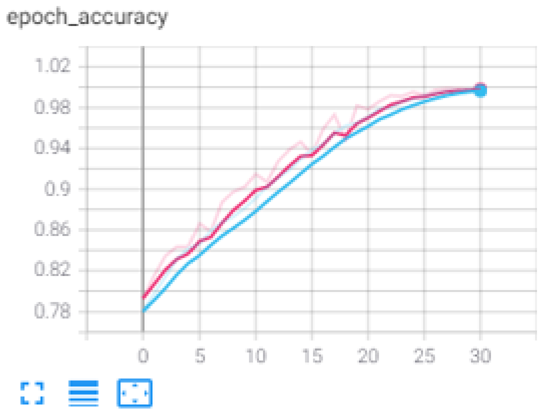
In [62]:

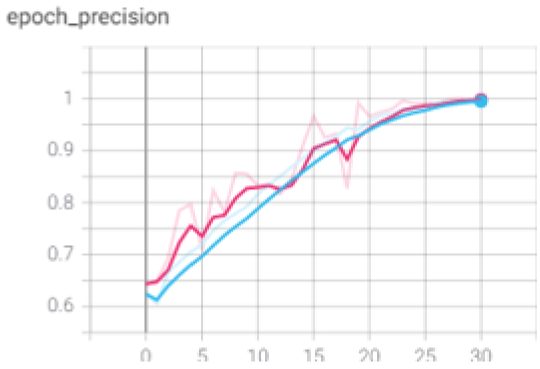
```
model1.evaluate(test_dataset_new)
```

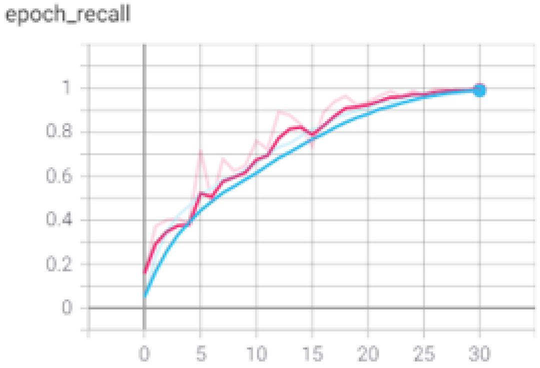
```
37/37 [==============================] - 92s 1s/step - loss: 0.0134 - accura
cy: 0.9992 - precision: 0.9962 - recall: 1.0000
```

Out[62]:

```
[0.013419254682958126, 0.9991554021835327, 0.9962049126625061, 1.0]
```







Confusion Matrix on Train Data with different threshold

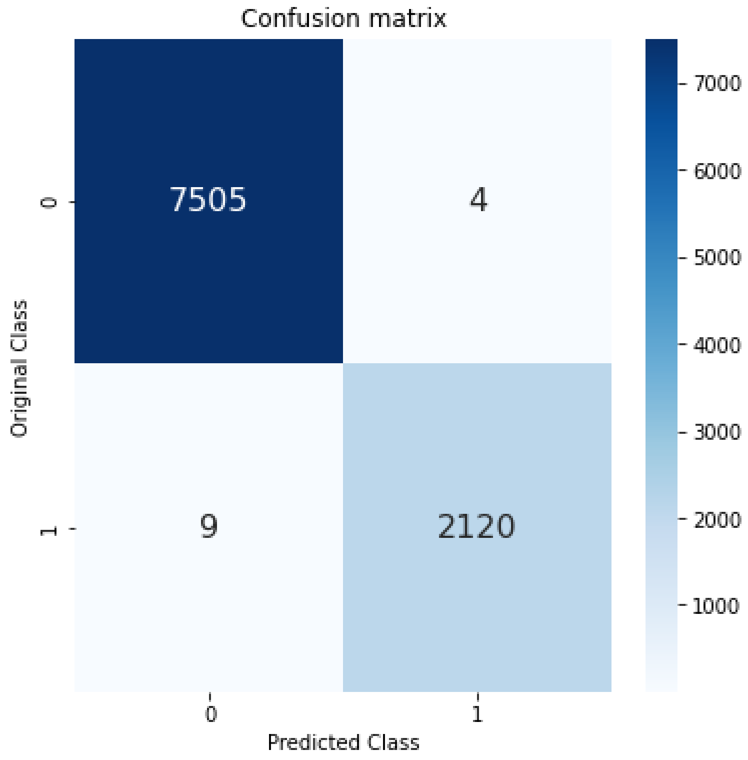
In [80]:

```
os.chdir('/gdrive/MyDrive/Image_Segementation_CS2/')
import seaborn as sns
from sklearn.metrics import confusion_matrix
def confusion_mat(test_y,predict_y):
  ''' Function to Visualize the Confusion Matrix'''
  labels = [0,1]
  plt.figure(figsize=(6,6))
  cmap=sns.light_palette("blue")
  C = confusion_matrix(test_y, predict_y)
  print("Percentage of misclassified points ",(len(test_y)-np.trace(C))/len(test_y)*100)
  sns.heatmap(C, cmap="Blues",annot=True,annot_kws={"size": 16},fmt='g')
  plt.xlabel('Predicted Class')
  plt.ylabel('Original Class')
  plt.title('Confusion matrix')
  plt.show()
```

In [83]:

```
y_pred_1=[] #array to store predicted label
y_true=[] #array to store the ground truth
for i,j in tqdm(train_dataset_new.take(9638)):

  y_pred_1.extend(model1.predict(i)) #predicting each batch
  y_true.extend(j)
  y_pred=[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.5: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

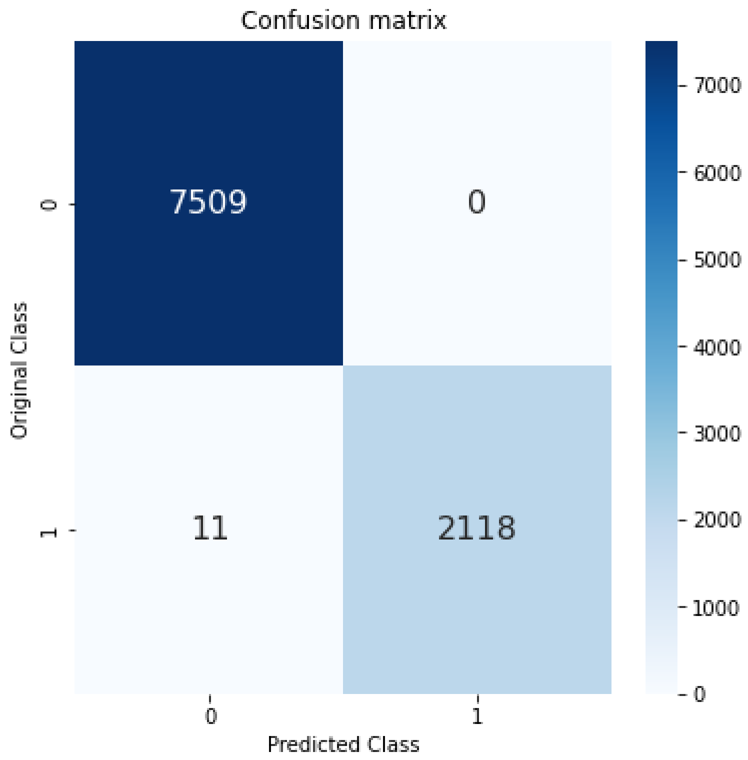
```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: TqdmDeprecat
ionWarning: This function will be removed in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
  This is separate from the ipykernel package so we can avoid doing imports
until

HBox(children=(FloatProgress(value=0.0, max=151.0), HTML(value='')))


Percentage of misclassified points  0.13488275575845612
```



In [87]:

```
y_pred =[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.6: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

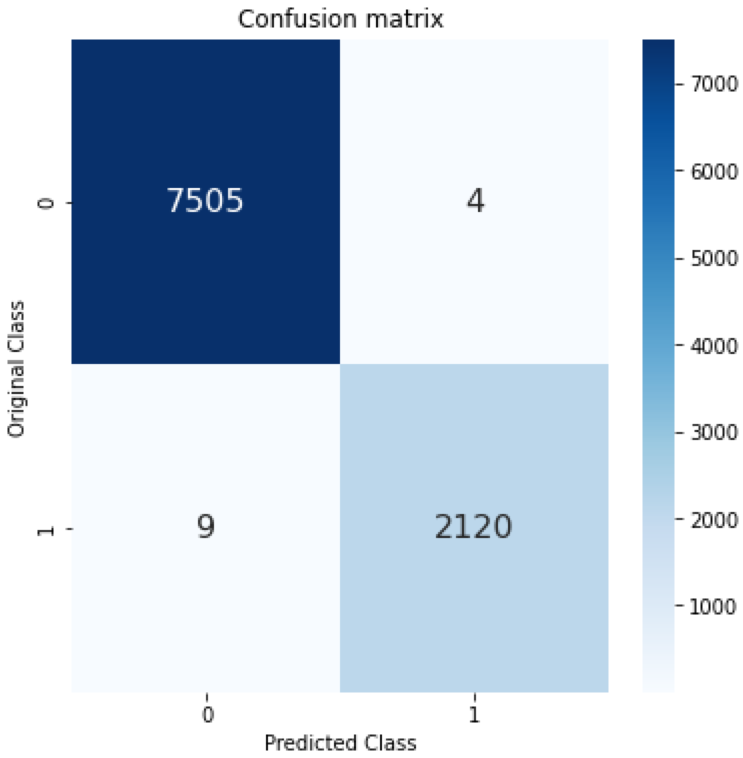
```
Percentage of misclassified points  0.11413156256484748
```



In [104]:

```
y_pred =[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.5: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

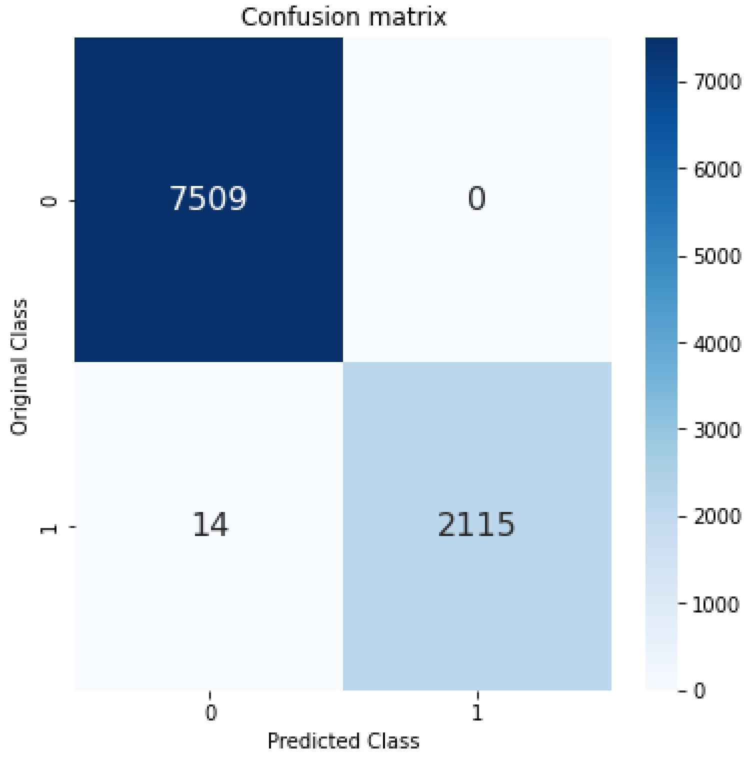
```
Percentage of misclassified points  0.13488275575845612
```



In [93]:

```
y_pred =[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.6: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

```
Percentage of misclassified points  0.1452583523552604
```

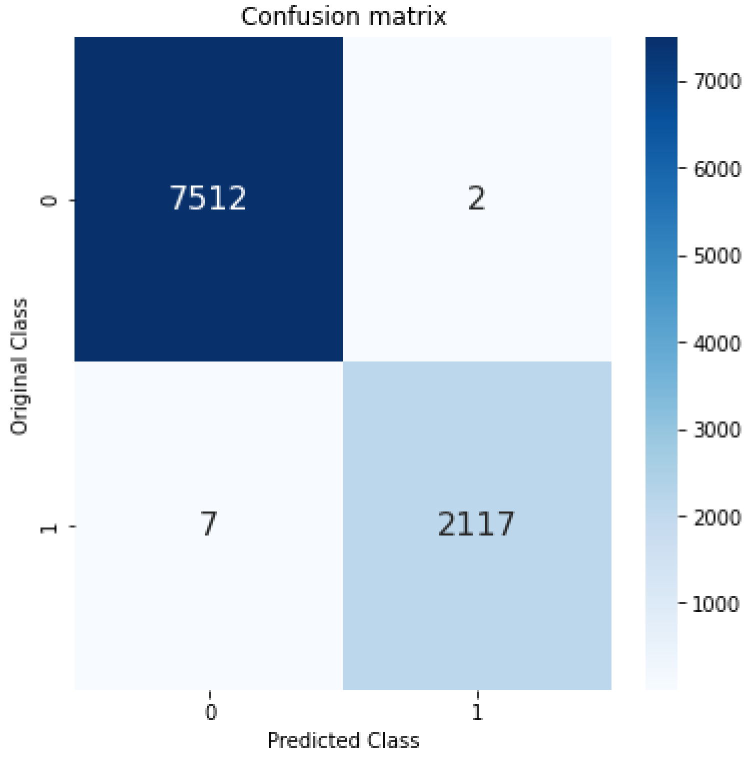


Confusion Matrix on test Data with different threshold

In [106]:

```
y_pred_1=[] #array to store predicted label
y_true=[] #array to store the ground truth
for i,j in tqdm(train_dataset_new.take(2129)):

  y_pred_1.extend(model1.predict(i)) #predicting each batch
  y_true.extend(j)
  y_pred=[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.5: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

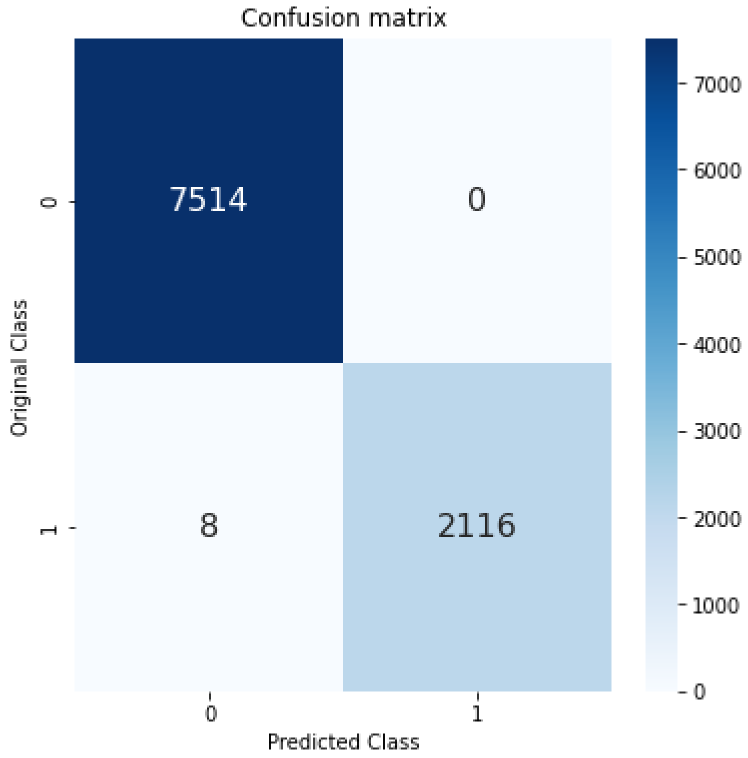
```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: TqdmDeprecat
ionWarning: This function will be removed in tqdm==5.0.0
Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
  This is separate from the ipykernel package so we can avoid doing imports
until

HBox(children=(FloatProgress(value=0.0, max=151.0), HTML(value='')))


Percentage of misclassified points  0.09338036937123885
```



In [107]:

```
y_pred =[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.6: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

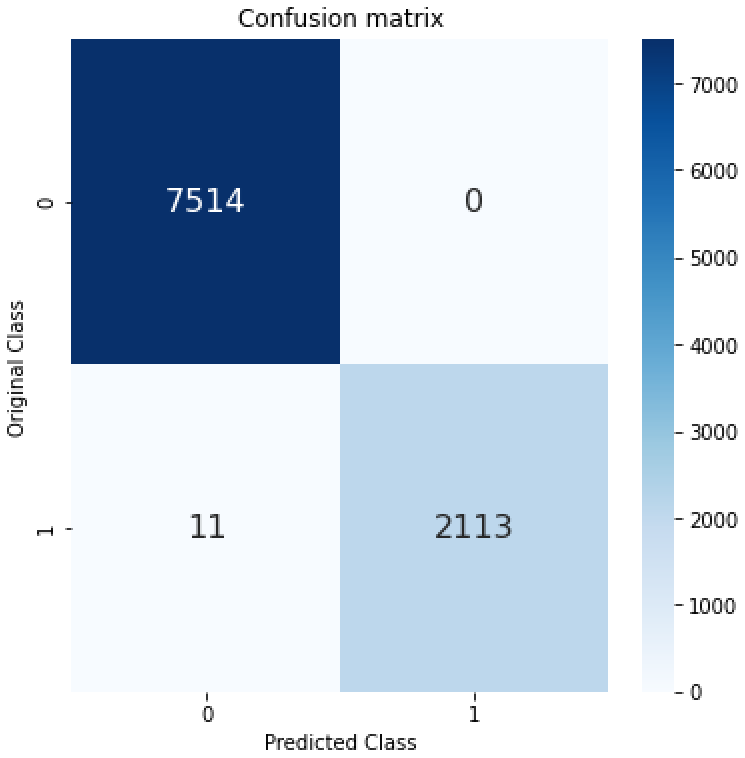
```
Percentage of misclassified points  0.08300477277443453
```



In [108]:

```
y_pred =[]
for i in y_pred_1: #the values are in probabilities and hence we are going to classify base
  if i[0]>=0.65: #setting threshold
    y_pred.append(1)
  else:
    y_pred.append(0)
y_true=np.array(y_true) #converting the array for into numpy
y_pred=np.array(y_pred).reshape(1,-1)
y_pred=y_pred[0]
confusion_mat(y_true,y_pred)
```

```
Percentage of misclassified points  0.11413156256484748
```



Observation

1. 0.5 as the threshold is better than the other thresholds for reducing False negatives
2. As the threshold increases the number of False negatives increases