In [42]:

```
#This file contains the neural network modelling of readmission after disch
arge.
#The first chunk is basic preprocessing
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn
import nltk
from nltk import word_tokenize
from nltk.probability import FreqDist
import string
from sklearn.metrics import roc_auc_score
from nltk.corpus import stopwords
from nltk.stem.snowball import SnowballStemmer
from sklearn.feature_extraction.text import CountVectorizer
import matplotlib.pyplot as plt
from sklearn.utils import class_weight
import tensorflow as tf
from keras.models import Sequential, load_model
from keras import optimizers
from keras.layers import Dense, Dropout, Conv1D, MaxPooling1D, Flatten, E
mbedding, LSTM
from keras.callbacks import EarlyStopping
from keras.callbacks import ModelCheckpoint
from keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.optimizers import Adam
import tensorflow as tf
from keras import backend as K
from sklearn.metrics import confusion_matrix
import seaborn as sns
from sklearn.metrics import accuracy_score
from sklearn.metrics import roc_curve
from sklearn.metrics import auc
from sklearn.metrics import roc_auc_score


#check for versions of TF
print("TF Version: ", tf.__version__)
print("Eager mode enabled: ", tf.executing_eagerly())
device_name = tf.test.gpu_device_name()
if device_name != '/device:GPU:0':
```

```
  raise SystemError('GPU device not found')
print('Found GPU at: {}'.format(device_name))
```

```
TF Version:  2.6.0
Eager mode enabled:  True
Found GPU at: /device:GPU:0
```

```
2021-10-23 20:37:18.882874: I tensorflow/stream_executor/cuda/cuda_gpu
_executor.cc:937] successful NUMA node read from SysFS had negative va
lue (-1), but there must be at least one NUMA node, so returning NUMA
node zero
2021-10-23 20:37:18.886891: I tensorflow/stream_executor/cuda/cuda_gpu
_executor.cc:937] successful NUMA node read from SysFS had negative va
lue (-1), but there must be at least one NUMA node, so returning NUMA
node zero
2021-10-23 20:37:18.887469: I tensorflow/stream_executor/cuda/cuda_gpu
_executor.cc:937] successful NUMA node read from SysFS had negative va
lue (-1), but there must be at least one NUMA node, so returning NUMA
node zero
2021-10-23 20:37:18.888003: I tensorflow/stream_executor/cuda/cuda_gpu
_executor.cc:937] successful NUMA node read from SysFS had negative va
lue (-1), but there must be at least one NUMA node, so returning NUMA
node zero
2021-10-23 20:37:18.888439: I tensorflow/stream_executor/cuda/cuda_gpu
_executor.cc:937] successful NUMA node read from SysFS had negative va
lue (-1), but there must be at least one NUMA node, so returning NUMA
node zero
2021-10-23 20:37:18.888801: I tensorflow/core/common_runtime/gpu/gpu_d
evice.cc:1510] Created device /device:GPU:0 with 15403 MB memory:  ->
device: 0, name: Tesla P100-PCIE-16GB, pci bus id: 0000:00:04.0, compu
te capability: 6.0
```

In [2]:

```
nltk.download('punkt')
df_admits = pd.read_csv('../input/admissions/ADMISSIONS.csv')
df_notes = pd.read_csv('../input/admissions/NOTEEVENTS.csv')
df_admits.ADMITTIME = pd.to_datetime(df_admits.ADMITTIME, format = '%Y-%m
-%d %H:%M:%S', errors = 'coerce')
df_admits.DISCHTIME = pd.to_datetime(df_admits.DISCHTIME, format = '%Y-%m
-%d %H:%M:%S', errors = 'coerce')
df_admits.DEATHTIME = pd.to_datetime(df_admits.DEATHTIME, format = '%Y-%m
-%d %H:%M:%S', errors = 'coerce')
df_admits = df_admits.sort_values(['SUBJECT_ID','ADMITTIME'])
df_admits = df_admits.reset_index(drop = True)
df_admits.columns
```

```
[nltk_data] Downloading package punkt to /usr/share/nltk_data...
[nltk_data]   Package punkt is already up-to-date!


/opt/conda/lib/python3.7/site-packages/IPython/core/interactiveshell.p
y:3441: DtypeWarning: Columns (4,5) have mixed types.Specify dtype opt
ion on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

Out[2]:

```
Index(['ROW_ID', 'SUBJECT_ID', 'HADM_ID', 'ADMITTIME', 'DISCHTIME',
       'DEATHTIME', 'ADMISSION_TYPE', 'ADMISSION_LOCATION',
       'DISCHARGE_LOCATION', 'INSURANCE', 'LANGUAGE', 'RELIGION',
       'MARITAL_STATUS', 'ETHNICITY', 'EDREGTIME', 'EDOUTTIME', 'DIAGN
OSIS',
       'HOSPITAL_EXPIRE_FLAG', 'HAS_CHARTEVENTS_DATA'],
      dtype='object')
```

In [3]:

```
##Create admission time variable, to be used to create binary target variable later
df_admits['NEXT_ADMIT'] = df_admits.groupby('SUBJECT_ID').ADMITTIME.shift(-1)
df_admits['NEXT_TYPE'] = df_admits.groupby('SUBJECT_ID').ADMISSION_TYPE.shift(-1)
df_admits['NEXT_ADMIT'].quantile(np.arange(0,1,.1))
```

Out[3]:

```
0.0   2100-08-27 11:37:00
0.1   2113-04-29 09:04:00
0.2   2122-12-22 16:33:00
0.3   2132-07-27 22:21:30
0.4   2142-02-15 02:27:00
0.5   2151-10-15 18:23:00
0.6   2162-01-18 13:00:00
0.7   2172-07-27 15:57:30
0.8   2182-07-15 15:12:00
0.9   2193-04-09 15:37:00
Name: NEXT_ADMIT, dtype: datetime64[ns]
```

In [4]:

```
#Do not use elective readmissions
rows = df_admits.NEXT_TYPE == 'ELECTIVE'
df_admits.loc[rows,'NEXT_ADMIT'] = pd.NaT
df_admits.loc[rows,'NEXT_TYPE'] = np.NaN
df_admits = df_admits.sort_values(['SUBJECT_ID','ADMITTIME'])
df_admits[['NEXT_ADMIT','NEXT_TYPE']] = df_admits.groupby(['SUBJECT_ID'])
[['NEXT_ADMIT','NEXT_TYPE']].fillna(method = 'bfill')
df_admits['DAYS']=  (df_admits.NEXT_ADMIT - df_admits.DISCHTIME).dt.total
_seconds()/(24*60*60)
df_admits['DAYS'].describe()
```

Out[4]:

```
count    11399.000000
mean       409.239700
std        639.190363
min        -18.765278
25%         23.976389
50%        120.199306
75%        507.237847
max       4107.968750
Name: DAYS, dtype: float64
```

In [5]:

```
#Only utilize discharge notes
df_notes_dis = df_notes.loc[df_notes.CATEGORY == 'Discharge summary']
df_notes_last = (df_notes_dis.groupby(['SUBJECT_ID','HADM_ID']).nth(-1)).
reset_index()
df_notes_last
```

Out[5]:

	SUBJECT_ID	HADM_ID	ROW_ID	CHARTDATE	CHARTTIME	STORETIME	CATEGORY
0	3	145834.0	44005	2101-10-31	NaN	NaN	Discharge summary
1	4	185777.0	4788	2191-03-23	NaN	NaN	Discharge summary
2	6	107064.0	20825	2175-06-15	NaN	NaN	Discharge summary
3	9	150750.0	57115	2149-11-14	NaN	NaN	Discharge summary
4	10	184167.0	17390	2103-07-06	NaN	NaN	Discharge summary
...	...	...	...	...	...	...	...
52721	99985	176670.0	51770	2181-02-12	NaN	NaN	Discharge summary
52722	99991	151118.0	9682	2185-01-05	NaN	NaN	Discharge summary
52723	99992	197084.0	41993	2144-07-28	NaN	NaN	Discharge summary
52724	99995	137810.0	42710	2147-02-11	NaN	NaN	Discharge summary
52725	99999	113369.0	52180	2118-01-04	NaN	NaN	Discharge summary

52726 rows × 11 columns

In [6]:

```
#Merge admissions and notes charts
df_adnotes = pd.merge(df_admits[['SUBJECT_ID','HADM_ID','ADMITTIME','DISC
HTIME','DAYS','NEXT_ADMIT','ADMISSION_TYPE','DEATHTIME']],
                      df_notes_last[['SUBJECT_ID','HADM_ID','TEXT']],
                      on = ['SUBJECT_ID','HADM_ID'],
                      how = 'left')
df_adnotes.groupby('ADMISSION_TYPE').apply(lambda g: g.TEXT.isnull().sum
())/df_adnotes.groupby('ADMISSION_TYPE').size()
df_adnotes
```

Out[6]:

	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DAYS	NEXT_ADMIT	ADMISSION_TY
0	2	163353	2138-07-17 19:04:00	2138-07-21 15:48:00	NaN	NaT	NEWBORN
1	3	145834	2101-10-20 19:08:00	2101-10-31 13:58:00	NaN	NaT	EMERGENCY
2	4	185777	2191-03-16 00:28:00	2191-03-23 18:41:00	NaN	NaT	EMERGENCY
3	5	178980	2103-02-02 04:31:00	2103-02-04 12:15:00	NaN	NaT	NEWBORN
4	6	107064	2175-05-30 07:15:00	2175-06-15 16:00:00	NaN	NaT	ELECTIVE
...	...	...	...	...	...	...	...
58971	99985	176670	2181-01-27 02:47:00	2181-02-12 17:05:00	NaN	NaT	EMERGENCY
58972	99991	151118	2184-12-24 08:30:00	2185-01-05 12:15:00	NaN	NaT	ELECTIVE
58973	99992	197084	2144-07-25 18:03:00	2144-07-28 17:56:00	NaN	NaT	EMERGENCY
58974	99995	137810	2147-02-08 08:00:00	2147-02-11 13:15:00	NaN	NaT	ELECTIVE
58975	99999	113369	2117-12-30 07:15:00	2118-01-04 16:30:00	NaN	NaT	ELECTIVE

58976 rows × 9 columns

In [7]:

```
#Create Target variable (whether a readmission occured within 30 days of di
scharge)
df_adnotes['OUTPUT_LABEL'] = (df_adnotes.DAYS < 30).astype('int')
df_adnotes = df_adnotes.sample(n = len(df_adnotes), random_state = 42)
df_adnotes = df_adnotes.reset_index(drop = True)
df_adnotes['OUTPUT_LABEL'].describe()
```

Out[7]:

```
count    58976.000000
mean         0.054717
std          0.227429
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          1.000000
Name: OUTPUT_LABEL, dtype: float64
```

In [8]:

```
#Only analyze patients who did not die in hospital
no_death = df_adnotes.DEATHTIME.isnull()
df_not_death = df_adnotes.loc[no_death].copy()
df_not_death = df_not_death.sample(n = len(df_not_death), random_state =
42)
df_not_death = df_not_death.reset_index(drop = True)
```

In [9]:

```
#Training Validation Test Split
df_valid_test=df_not_death.sample(frac=0.20,random_state=42)
df_test = df_valid_test.sample(frac = 0.5, random_state = 42)
df_valid = df_valid_test.drop(df_test.index)
df_train_all=df_not_death.drop(df_valid_test.index)
rows_pos = df_train_all.OUTPUT_LABEL == 1
df_train_pos = df_train_all.loc[rows_pos]
df_train_neg = df_train_all.loc[~rows_pos]
#Use undersampling of negative cases
df_train = pd.concat([df_train_pos, df_train_neg.sample(n = len(df_train_
pos), random_state = 42)],axis = 0)
df_train = df_train.sample(n = len(df_train), random_state = 42).reset_in
dex(drop = True)
```

In [10]:

```
#Preprocess text
def preprocess_text(df):
    df.TEXT = df.TEXT.fillna(' ')
    df.TEXT = df.TEXT.str.replace('\n',' ')
    df.TEXT = df.TEXT.str.replace('\r',' ')
    df.TEXT = df.TEXT.str.replace('[^A-Za-z0-9(),!?@\'\`\"\_\n]', ' ')
    return df
df_train = preprocess_text(df_train)
df_test = preprocess_text(df_test)
df_valid = preprocess_text(df_valid)
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:6: Future
Warning: The default value of regex will change from True to False in
a future version.
```

In [11]:
```
#Remove stopwords, stem
sw = ['the','and','to','of','was','with','a','on','in','for','name',
      'is','patient','s','he','at','as','or','one','she','his','her','am'
,
      'were','you','pt','pm','by','be','had','your','this','date',
      'from','there','an','that','p','are','have','has','h','but','o',
      'namepattern','which','every','also','should','if','it','been','wh
o','during', 'x']
stemmer = SnowballStemmer("english")
def stemming(text):
    text = [stemmer.stem(word) for word in text.split()]
    return " ".join(text)
df_train['TEXT'] = df_train['TEXT'].apply(stemming)
df_test['TEXT'] = df_test['TEXT'].apply(stemming)
df_valid['TEXT'] = df_valid['TEXT'].apply(stemming)
def tokenizer_better(text):
    punc_list = string.punctuation+'0123456789'
    t = str.maketrans(dict.fromkeys(punc_list, " "))
    text = text.lower().translate(t)
    tokens = word_tokenize(text)
    return tokens
```

In [12]:
```
#Vectorize and fit training and test text
vect = CountVectorizer(tokenizer = tokenizer_better, stop_words =sw, min_
df = 5, max_df =.9,)
vect.fit(df_train.TEXT.values.astype('U'))
dictionary = vect.vocabulary_.items()
X_train_tf = vect.transform(df_train.TEXT.values.astype('U'))
X_test_tf = vect.transform(df_test.TEXT.values.astype('U'))
X_valid_tf = vect.transform(df_valid.TEXT.values.astype('U'))
y_train = df_train.OUTPUT_LABEL
y_test = df_test.OUTPUT_LABEL
y_valid = df_valid.OUTPUT_LABEL
```

```
/opt/conda/lib/python3.7/site-packages/sklearn/feature_extraction/tex
t.py:484: UserWarning: The parameter 'token_pattern' will not be used
since 'tokenizer' is not None'
  warnings.warn("The parameter 'token_pattern' will not be used"
```

In [13]:
```
y_train
```

Out[13]:
```
0       1
1       0
2       1
3       1
4       1
       ..
4999    0
5000    1
5001    0
5002    0
5003    1
Name: OUTPUT_LABEL, Length: 5004, dtype: int64
```

In [34]:
```
plt.style.use('ggplot')
def plot_history(history):
    acc = history.history['my_auc']
    val_acc = history.history['val_my_auc']
    loss = history.history['loss']
    val_loss = history.history['val_loss']
    x = range(1, len(acc) + 1)

    plt.figure(figsize=(12, 5))
    plt.subplot(1, 2, 1)
    plt.plot(x, acc, 'b', label='Training auc')
    plt.plot(x, val_acc, 'r', label='Validation auc')
    plt.title('Training and validation AUC')
    plt.legend()
    plt.subplot(1, 2, 2)
    plt.plot(x, loss, 'b', label='Training loss')
    plt.plot(x, val_loss, 'r', label='Validation loss')
    plt.title('Training and validation loss')
    plt.legend()
```

In [39]:

```
with tf.device('/device:GPU:0'):
    model = Sequential()
    model.add(Dense(units=256, activation='relu', input_dim=len(vect.get_
feature_names())))
    model.add(Dropout(0.2))
    model.add(Dense(128, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(64, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(64, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(units=1, activation='sigmoid'))
    Adam = tf.keras.optimizers.Adam(lr=0.00001)
    model.compile(loss = 'binary_crossentropy', optimizer = Adam, metrics
=[tf.keras.metrics.AUC(name='my_auc')])
    model.summary()
    es = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patienc
e=25)
    mc = ModelCheckpoint('best_model.h5', monitor='val_loss', mode='min',
verbose=1, save_best_only=True)
    history = model.fit(X_train_tf.toarray(), y_train,epochs=500, batch_s
ize=128, verbose=1,validation_data=(X_valid_tf.toarray(), y_valid),callba
cks=[es,mc])
    saved_model = load_model('best_model.h5')
    scores = saved_model.evaluate(X_test_tf.toarray(), y_test, verbose=1)
    print("AUC:", scores[1])
    plot_history(history)
```

```
Model: "sequential_9"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_45 (Dense)             (None, 256)               3238144
_________________________________________________________________
dropout_36 (Dropout)         (None, 256)               0
_________________________________________________________________
dense_46 (Dense)             (None, 128)               32896
_________________________________________________________________
dropout_37 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_47 (Dense)             (None, 64)                8256
_________________________________________________________________
dropout_38 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_48 (Dense)             (None, 64)                4160
_________________________________________________________________
dropout_39 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_49 (Dense)             (None, 1)                 65
=================================================================
Total params: 3,283,521
Trainable params: 3,283,521
Non-trainable params: 0
_________________________________________________________________
Epoch 1/500
40/40 [==============================] - 2s 46ms/step - loss: 0.6998 -
my_auc: 0.5453 - val_loss: 0.7428 - val_my_auc: 0.5898

Epoch 00001: val_loss improved from inf to 0.74277, saving model to be
st_model.h5
Epoch 2/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6961 -
my_auc: 0.5533 - val_loss: 0.7397 - val_my_auc: 0.6136

Epoch 00002: val_loss improved from 0.74277 to 0.73971, saving model t
o best_model.h5
Epoch 3/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6898 -
my_auc: 0.5750 - val_loss: 0.7421 - val_my_auc: 0.6319
```

```
Epoch 00003: val_loss did not improve from 0.73971
Epoch 4/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6894 -
my_auc: 0.5793 - val_loss: 0.7358 - val_my_auc: 0.6445

Epoch 00004: val_loss improved from 0.73971 to 0.73580, saving model t
o best_model.h5
Epoch 5/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6783 -
my_auc: 0.6122 - val_loss: 0.7353 - val_my_auc: 0.6507

Epoch 00005: val_loss improved from 0.73580 to 0.73532, saving model t
o best_model.h5
Epoch 6/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6756 -
my_auc: 0.6132 - val_loss: 0.7289 - val_my_auc: 0.6586

Epoch 00006: val_loss improved from 0.73532 to 0.72889, saving model t
o best_model.h5
Epoch 7/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6729 -
my_auc: 0.6230 - val_loss: 0.7305 - val_my_auc: 0.6654

Epoch 00007: val_loss did not improve from 0.72889
Epoch 8/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6692 -
my_auc: 0.6389 - val_loss: 0.7242 - val_my_auc: 0.6686

Epoch 00008: val_loss improved from 0.72889 to 0.72416, saving model t
o best_model.h5
Epoch 9/500
40/40 [==============================] - 1s 20ms/step - loss: 0.6689 -
my_auc: 0.6398 - val_loss: 0.7237 - val_my_auc: 0.6730

Epoch 00009: val_loss improved from 0.72416 to 0.72371, saving model t
o best_model.h5
Epoch 10/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6643 -
my_auc: 0.6470 - val_loss: 0.7229 - val_my_auc: 0.6772

Epoch 00010: val_loss improved from 0.72371 to 0.72294, saving model t
```

```
o best_model.h5
Epoch 11/500
40/40 [==============================] - 1s 18ms/step - loss: 0.6585 -
my_auc: 0.6631 - val_loss: 0.7107 - val_my_auc: 0.6789

Epoch 00011: val_loss improved from 0.72294 to 0.71072, saving model t
o best_model.h5
Epoch 12/500
40/40 [==============================] - 1s 24ms/step - loss: 0.6571 -
my_auc: 0.6667 - val_loss: 0.7107 - val_my_auc: 0.6818

Epoch 00012: val_loss improved from 0.71072 to 0.71067, saving model t
o best_model.h5
Epoch 13/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6533 -
my_auc: 0.6713 - val_loss: 0.7146 - val_my_auc: 0.6841

Epoch 00013: val_loss did not improve from 0.71067
Epoch 14/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6482 -
my_auc: 0.6807 - val_loss: 0.7109 - val_my_auc: 0.6872

Epoch 00014: val_loss did not improve from 0.71067
Epoch 15/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6479 -
my_auc: 0.6847 - val_loss: 0.7037 - val_my_auc: 0.6891

Epoch 00015: val_loss improved from 0.71067 to 0.70369, saving model t
o best_model.h5
Epoch 16/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6400 -
my_auc: 0.6960 - val_loss: 0.6996 - val_my_auc: 0.6891

Epoch 00016: val_loss improved from 0.70369 to 0.69959, saving model t
o best_model.h5
Epoch 17/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6379 -
my_auc: 0.7046 - val_loss: 0.6936 - val_my_auc: 0.6915

Epoch 00017: val_loss improved from 0.69959 to 0.69364, saving model t
o best_model.h5
Epoch 18/500
```

```
40/40 [==============================] - 1s 16ms/step - loss: 0.6335 -
my_auc: 0.7052 - val_loss: 0.6812 - val_my_auc: 0.6924

Epoch 00018: val_loss improved from 0.69364 to 0.68115, saving model t
o best_model.h5
Epoch 19/500
40/40 [==============================] - 1s 15ms/step - loss: 0.6288 -
my_auc: 0.7143 - val_loss: 0.6838 - val_my_auc: 0.6928

Epoch 00019: val_loss did not improve from 0.68115
Epoch 20/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6261 -
my_auc: 0.7154 - val_loss: 0.6928 - val_my_auc: 0.6946

Epoch 00020: val_loss did not improve from 0.68115
Epoch 21/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6211 -
my_auc: 0.7224 - val_loss: 0.6883 - val_my_auc: 0.6956

Epoch 00021: val_loss did not improve from 0.68115
Epoch 22/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6166 -
my_auc: 0.7277 - val_loss: 0.6877 - val_my_auc: 0.6961

Epoch 00022: val_loss did not improve from 0.68115
Epoch 23/500
40/40 [==============================] - 1s 19ms/step - loss: 0.6128 -
my_auc: 0.7355 - val_loss: 0.6762 - val_my_auc: 0.6972

Epoch 00023: val_loss improved from 0.68115 to 0.67623, saving model t
o best_model.h5
Epoch 24/500
40/40 [==============================] - 1s 17ms/step - loss: 0.6085 -
my_auc: 0.7377 - val_loss: 0.6772 - val_my_auc: 0.6982

Epoch 00024: val_loss did not improve from 0.67623
Epoch 25/500
40/40 [==============================] - 1s 16ms/step - loss: 0.6031 -
my_auc: 0.7483 - val_loss: 0.6892 - val_my_auc: 0.6984

Epoch 00025: val_loss did not improve from 0.67623
Epoch 26/500
```

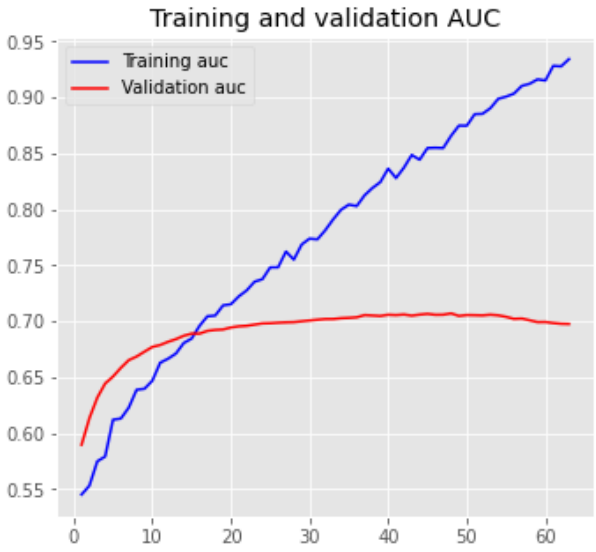
```
40/40 [==============================] - 1s 16ms/step - loss: 0.6002 -
my_auc: 0.7484 - val_loss: 0.6766 - val_my_auc: 0.6988

Epoch 00026: val_loss did not improve from 0.67623
Epoch 27/500
40/40 [==============================] - 1s 17ms/step - loss: 0.5913 -
my_auc: 0.7624 - val_loss: 0.6707 - val_my_auc: 0.6992

Epoch 00027: val_loss improved from 0.67623 to 0.67074, saving model t
o best_model.h5
Epoch 28/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5949 -
my_auc: 0.7552 - val_loss: 0.6763 - val_my_auc: 0.6993

Epoch 00028: val_loss did not improve from 0.67074
Epoch 29/500
40/40 [==============================] - 1s 17ms/step - loss: 0.5872 -
my_auc: 0.7688 - val_loss: 0.6711 - val_my_auc: 0.7001

Epoch 00029: val_loss did not improve from 0.67074
Epoch 30/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5819 -
my_auc: 0.7740 - val_loss: 0.6724 - val_my_auc: 0.7009

Epoch 00030: val_loss did not improve from 0.67074
Epoch 31/500
40/40 [==============================] - 1s 17ms/step - loss: 0.5783 -
my_auc: 0.7733 - val_loss: 0.6747 - val_my_auc: 0.7015

Epoch 00031: val_loss did not improve from 0.67074
Epoch 32/500
40/40 [==============================] - 1s 17ms/step - loss: 0.5727 -
my_auc: 0.7815 - val_loss: 0.6754 - val_my_auc: 0.7021

Epoch 00032: val_loss did not improve from 0.67074
Epoch 33/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5638 -
my_auc: 0.7912 - val_loss: 0.6726 - val_my_auc: 0.7021

Epoch 00033: val_loss did not improve from 0.67074
Epoch 34/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5581 -
```

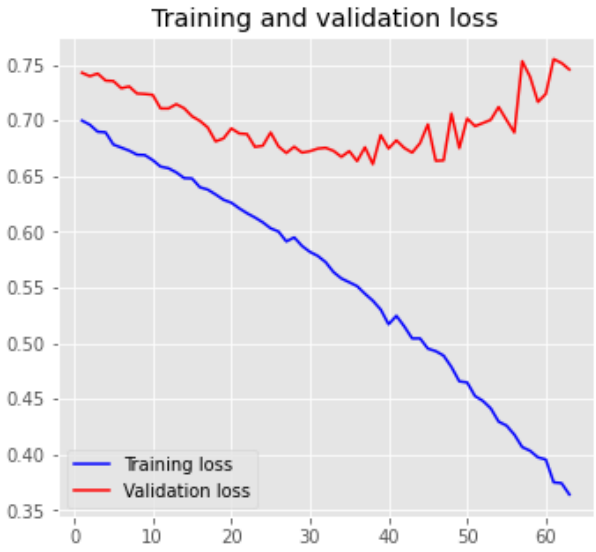
```
my_auc: 0.7997 - val_loss: 0.6673 - val_my_auc: 0.7029

Epoch 00034: val_loss improved from 0.67074 to 0.66726, saving model t
o best_model.h5
Epoch 35/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5547 -
my_auc: 0.8044 - val_loss: 0.6724 - val_my_auc: 0.7032

Epoch 00035: val_loss did not improve from 0.66726
Epoch 36/500
40/40 [==============================] - 1s 15ms/step - loss: 0.5512 -
my_auc: 0.8030 - val_loss: 0.6634 - val_my_auc: 0.7037

Epoch 00036: val_loss improved from 0.66726 to 0.66342, saving model t
o best_model.h5
Epoch 37/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5444 -
my_auc: 0.8127 - val_loss: 0.6759 - val_my_auc: 0.7056

Epoch 00037: val_loss did not improve from 0.66342
Epoch 38/500
40/40 [==============================] - 1s 15ms/step - loss: 0.5381 -
my_auc: 0.8192 - val_loss: 0.6608 - val_my_auc: 0.7053

Epoch 00038: val_loss improved from 0.66342 to 0.66075, saving model t
o best_model.h5
Epoch 39/500
40/40 [==============================] - 1s 22ms/step - loss: 0.5302 -
my_auc: 0.8244 - val_loss: 0.6868 - val_my_auc: 0.7049

Epoch 00039: val_loss did not improve from 0.66075
Epoch 40/500
40/40 [==============================] - 1s 18ms/step - loss: 0.5169 -
my_auc: 0.8365 - val_loss: 0.6748 - val_my_auc: 0.7060

Epoch 00040: val_loss did not improve from 0.66075
Epoch 41/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5245 -
my_auc: 0.8281 - val_loss: 0.6821 - val_my_auc: 0.7056

Epoch 00041: val_loss did not improve from 0.66075
Epoch 42/500
```

```
40/40 [==============================] - 1s 17ms/step - loss: 0.5150 -
my_auc: 0.8374 - val_loss: 0.6753 - val_my_auc: 0.7062

Epoch 00042: val_loss did not improve from 0.66075
Epoch 43/500
40/40 [==============================] - 1s 17ms/step - loss: 0.5042 -
my_auc: 0.8487 - val_loss: 0.6711 - val_my_auc: 0.7052

Epoch 00043: val_loss did not improve from 0.66075
Epoch 44/500
40/40 [==============================] - 1s 16ms/step - loss: 0.5043 -
my_auc: 0.8444 - val_loss: 0.6795 - val_my_auc: 0.7061

Epoch 00044: val_loss did not improve from 0.66075
Epoch 45/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4951 -
my_auc: 0.8549 - val_loss: 0.6965 - val_my_auc: 0.7067

Epoch 00045: val_loss did not improve from 0.66075
Epoch 46/500
40/40 [==============================] - 1s 17ms/step - loss: 0.4928 -
my_auc: 0.8552 - val_loss: 0.6637 - val_my_auc: 0.7059

Epoch 00046: val_loss did not improve from 0.66075
Epoch 47/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4888 -
my_auc: 0.8549 - val_loss: 0.6641 - val_my_auc: 0.7060

Epoch 00047: val_loss did not improve from 0.66075
Epoch 48/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4786 -
my_auc: 0.8658 - val_loss: 0.7062 - val_my_auc: 0.7072

Epoch 00048: val_loss did not improve from 0.66075
Epoch 49/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4657 -
my_auc: 0.8749 - val_loss: 0.6752 - val_my_auc: 0.7049

Epoch 00049: val_loss did not improve from 0.66075
Epoch 50/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4647 -
my_auc: 0.8748 - val_loss: 0.7018 - val_my_auc: 0.7057
```

```
Epoch 00050: val_loss did not improve from 0.66075
Epoch 51/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4524 -
my_auc: 0.8851 - val_loss: 0.6948 - val_my_auc: 0.7056

Epoch 00051: val_loss did not improve from 0.66075
Epoch 52/500
40/40 [==============================] - 1s 17ms/step - loss: 0.4482 -
my_auc: 0.8855 - val_loss: 0.6976 - val_my_auc: 0.7054

Epoch 00052: val_loss did not improve from 0.66075
Epoch 53/500
40/40 [==============================] - 1s 17ms/step - loss: 0.4414 -
my_auc: 0.8905 - val_loss: 0.7005 - val_my_auc: 0.7060

Epoch 00053: val_loss did not improve from 0.66075
Epoch 54/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4295 -
my_auc: 0.8989 - val_loss: 0.7121 - val_my_auc: 0.7054

Epoch 00054: val_loss did not improve from 0.66075
Epoch 55/500
40/40 [==============================] - 1s 20ms/step - loss: 0.4260 -
my_auc: 0.9008 - val_loss: 0.7004 - val_my_auc: 0.7041

Epoch 00055: val_loss did not improve from 0.66075
Epoch 56/500
40/40 [==============================] - 1s 18ms/step - loss: 0.4179 -
my_auc: 0.9035 - val_loss: 0.6891 - val_my_auc: 0.7023

Epoch 00056: val_loss did not improve from 0.66075
Epoch 57/500
40/40 [==============================] - 1s 16ms/step - loss: 0.4066 -
my_auc: 0.9104 - val_loss: 0.7532 - val_my_auc: 0.7026

Epoch 00057: val_loss did not improve from 0.66075
Epoch 58/500
40/40 [==============================] - 1s 20ms/step - loss: 0.4034 -
my_auc: 0.9123 - val_loss: 0.7392 - val_my_auc: 0.7010

Epoch 00058: val_loss did not improve from 0.66075
```

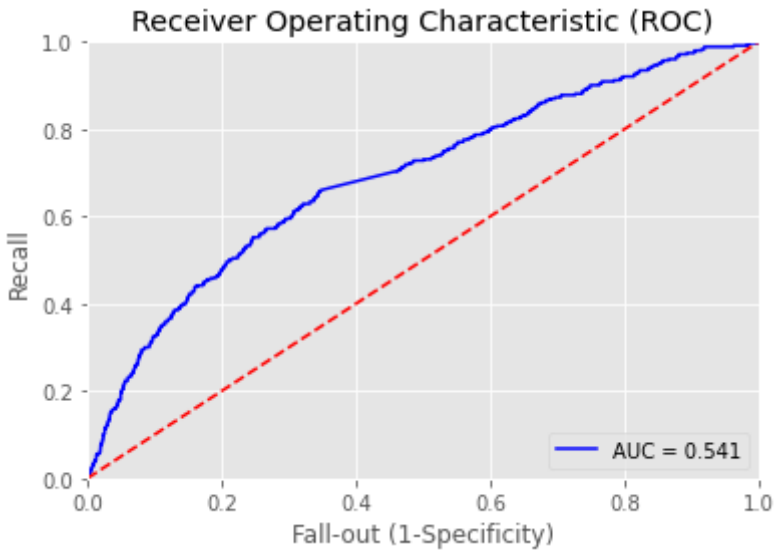
```
Epoch 59/500
40/40 [==============================] - 1s 23ms/step - loss: 0.3975 -
my_auc: 0.9163 - val_loss: 0.7165 - val_my_auc: 0.6994

Epoch 00059: val_loss did not improve from 0.66075
Epoch 60/500
40/40 [==============================] - 1s 16ms/step - loss: 0.3955 -
my_auc: 0.9153 - val_loss: 0.7238 - val_my_auc: 0.6995

Epoch 00060: val_loss did not improve from 0.66075
Epoch 61/500
40/40 [==============================] - 1s 17ms/step - loss: 0.3750 -
my_auc: 0.9283 - val_loss: 0.7551 - val_my_auc: 0.6985

Epoch 00061: val_loss did not improve from 0.66075
Epoch 62/500
40/40 [==============================] - 1s 17ms/step - loss: 0.3743 -
my_auc: 0.9279 - val_loss: 0.7515 - val_my_auc: 0.6979

Epoch 00062: val_loss did not improve from 0.66075
Epoch 63/500
40/40 [==============================] - 1s 17ms/step - loss: 0.3640 -
my_auc: 0.9344 - val_loss: 0.7457 - val_my_auc: 0.6977

Epoch 00063: val_loss did not improve from 0.66075
Epoch 00063: early stopping
166/166 [==============================] - 1s 3ms/step - loss: 0.6728
- my_auc: 0.6916
AUC: 0.6915538311004639
```

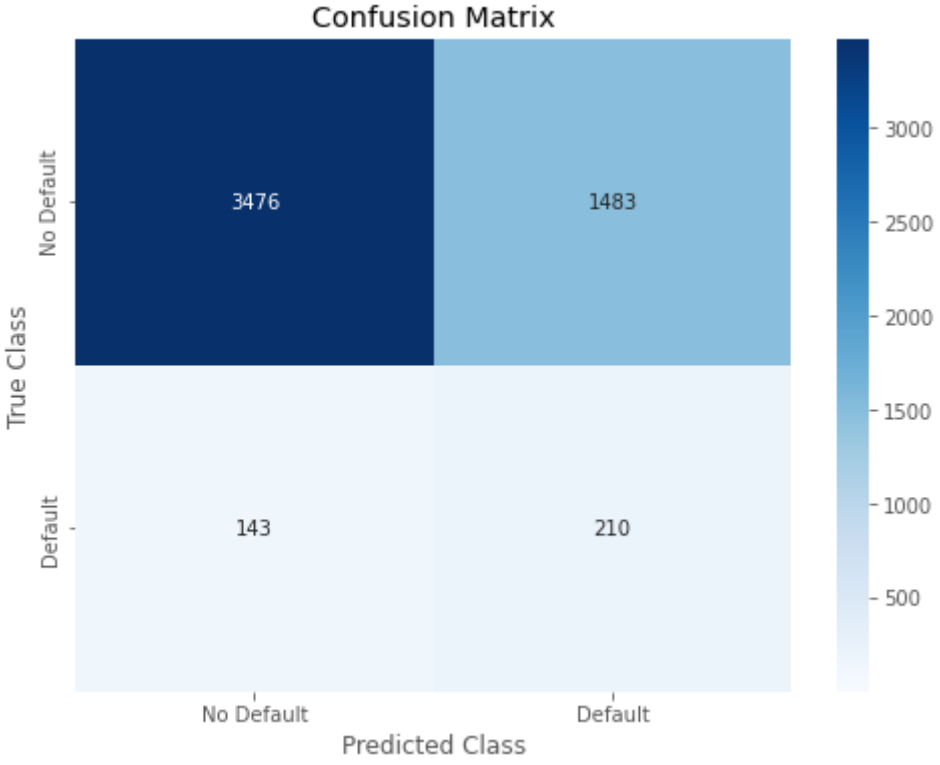

Training and validation AUC
Training auc
Validation auc
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
10
20
30
40
50
60


Training and validation loss
Training loss
Validation loss
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0
10
20
30
40
50
60

In [40]:

```
#predictions, confusion matrix, and ROC curve
predictions_NN_prob = saved_model.predict(X_test_tf)
predictions_NN_prob = predictions_NN_prob[:,0]
predictions_NN_01 = np.where (predictions_NN_prob >.5,1,0)
confusion_matrix(y_test, predictions_NN_01)
false_positive_rate, recall, thresholds = roc_curve(y_test, predictions_N
N_prob)
roc_auc = auc(false_positive_rate, recall)
plt.figure()
plt.title('Receiver Operating Characteristic (ROC)')
plt.plot(false_positive_rate, recall, 'b', label = 'AUC = %0.3f' %roc_auc
)
plt.legend(loc='lower right')
plt.plot([0,1], [0,1], 'r--')
plt.xlim([0.0,1.0])
plt.ylim([0.0,1.0])
plt.ylabel('Recall')
plt.xlabel('Fall-out (1-Specificity)')
plt.show()
cm = confusion_matrix(y_test, predictions_NN_01)
labels = ['No Default', 'Default']
plt.figure(figsize=(8,6))
sns.heatmap(cm,xticklabels=labels, yticklabels=labels, annot=True, fmt=
'd', cmap="Blues", vmin = 0.2);
plt.title('Confusion Matrix')
plt.ylabel('True Class')
plt.xlabel('Predicted Class')
plt.show()
```


Receiver Operating Characteristic (ROC)
Recall
Fall-out (1-Specificity)
AUC = 0.541
0.0
0.2
0.4
0.6
0.8
1.0


Confusion Matrix
True Class
No Default
Default
3476
1483
143
210
Predicted Class
3000
2500
2000
1500
1000
500

In [49]:

```
#Trying a Convolutional Neural Network with word sequences
with tf.device('/device:GPU:0'):
    word2idx = {word: idx for idx, word in enumerate(vect.get_feature_nam
es())}
    tokenize = vect.build_tokenizer()
    preprocess = vect.build_preprocessor()
    def to_sequence(tokenizer, preprocessor, index, text):
        words = tokenizer(preprocessor(text))
        indexes = [index[word] for word in words if word in index]
        return indexes
    X_train_sequences = [to_sequence(tokenize, preprocess, word2idx, x) f
or x in df_train.TEXT]
    MAX_SEQ_LENGTH = len(max(X_train_sequences, key=len))
    print("MAX_SEQ_LENGTH=", MAX_SEQ_LENGTH)
    N_FEATURES = len(vect.get_feature_names())
    X_train_sequences = pad_sequences(X_train_sequences, maxlen=MAX_SEQ_L
ENGTH, value=N_FEATURES)
    X_test_sequences = [to_sequence(tokenize, preprocess, word2idx, x) fo
r x in df_test.TEXT]
    X_test_sequences = pad_sequences(X_test_sequences, maxlen=MAX_SEQ_LEN
GTH, value=N_FEATURES)
    X_valid_sequences = [to_sequence(tokenize, preprocess, word2idx, x) f
or x in df_valid.TEXT]
    X_valid_sequences = pad_sequences(X_valid_sequences, maxlen=MAX_SEQ_L
ENGTH, value=N_FEATURES)
    print(X_train_sequences[0])
```

```
MAX_SEQ_LENGTH= 4832
[12648 12648 12648 ...  6743  7703  2232]
```

In [50]:

```
with tf.device('/device:GPU:0'):
    model = Sequential()
    model.add(Embedding(len(vect.get_feature_names()) + 1,
                    64,
                    input_length=MAX_SEQ_LENGTH))
    model.add(Conv1D(64, 5, activation='relu'))
    model.add(MaxPooling1D(5))
    model.add(Flatten())
    model.add(Dense(units=64, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(units=1, activation='sigmoid'))
    Adam = tf.keras.optimizers.Adam(lr=0.00001)
    model.compile(loss='binary_crossentropy', optimizer= Adam , metrics=[
tf.keras.metrics.AUC(name='my_auc')])
    print(model.summary())
    es = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patienc
e=25)
    mc = ModelCheckpoint('best_model_one.h5', monitor='val_loss', mode='m
in', verbose=1, save_best_only=True)
    hist = model.fit(X_train_sequences, y_train,
          epochs=500, batch_size=64, verbose=1,
          validation_data=(X_valid_sequences, y_valid),callbacks=[es,mc])
    saved_model_one = load_model('best_model_one.h5')
    scores = saved_model_one.evaluate(X_test_sequences, y_test, verbose=1
)
    print("AUC:", scores[1])
    plot_history(hist)
```

```
Model: "sequential_13"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_3 (Embedding)      (None, 4832, 64)          809536
_________________________________________________________________
conv1d_3 (Conv1D)            (None, 4828, 64)          20544
_________________________________________________________________
max_pooling1d_3 (MaxPooling1 (None, 965, 64)           0
_________________________________________________________________
flatten_3 (Flatten)          (None, 61760)             0
_________________________________________________________________
dense_56 (Dense)             (None, 64)                3952704
_________________________________________________________________
dropout_43 (Dropout)         (None, 64)                0
_________________________________________________________________
dense_57 (Dense)             (None, 1)                 65
=================================================================
Total params: 4,782,849
Trainable params: 4,782,849
Non-trainable params: 0
_________________________________________________________________
None
Epoch 1/500


2021-10-23 20:52:47.365856: I tensorflow/stream_executor/cuda/cuda_dn
n.cc:369] Loaded cuDNN version 8005
```

```
79/79 [==============================] - 9s 42ms/step - loss: 0.6836 -
my_auc: 0.5900 - val_loss: 0.7235 - val_my_auc: 0.6157

Epoch 00001: val_loss improved from inf to 0.72346, saving model to be
st_model_one.h5
Epoch 2/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6750 -
my_auc: 0.6110 - val_loss: 0.7040 - val_my_auc: 0.6213

Epoch 00002: val_loss improved from 0.72346 to 0.70405, saving model t
o best_model_one.h5
Epoch 3/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6726 -
my_auc: 0.6174 - val_loss: 0.6827 - val_my_auc: 0.6259

Epoch 00003: val_loss improved from 0.70405 to 0.68272, saving model t
o best_model_one.h5
Epoch 4/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6709 -
my_auc: 0.6207 - val_loss: 0.6956 - val_my_auc: 0.6290

Epoch 00004: val_loss did not improve from 0.68272
Epoch 5/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6703 -
my_auc: 0.6212 - val_loss: 0.7071 - val_my_auc: 0.6295

Epoch 00005: val_loss did not improve from 0.68272
Epoch 6/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6705 -
my_auc: 0.6197 - val_loss: 0.6694 - val_my_auc: 0.6302

Epoch 00006: val_loss improved from 0.68272 to 0.66943, saving model t
o best_model_one.h5
Epoch 7/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6687 -
my_auc: 0.6247 - val_loss: 0.6686 - val_my_auc: 0.6333

Epoch 00007: val_loss improved from 0.66943 to 0.66863, saving model t
o best_model_one.h5
Epoch 8/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6665 -
```

```
my_auc: 0.6317 - val_loss: 0.7180 - val_my_auc: 0.6326

Epoch 00008: val_loss did not improve from 0.66863
Epoch 9/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6661 -
my_auc: 0.6343 - val_loss: 0.6303 - val_my_auc: 0.6355

Epoch 00009: val_loss improved from 0.66863 to 0.63028, saving model t
o best_model_one.h5
Epoch 10/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6662 -
my_auc: 0.6319 - val_loss: 0.7131 - val_my_auc: 0.6338

Epoch 00010: val_loss did not improve from 0.63028
Epoch 11/500
79/79 [==============================] - 3s 38ms/step - loss: 0.6644 -
my_auc: 0.6373 - val_loss: 0.6915 - val_my_auc: 0.6347

Epoch 00011: val_loss did not improve from 0.63028
Epoch 12/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6638 -
my_auc: 0.6377 - val_loss: 0.7210 - val_my_auc: 0.6354

Epoch 00012: val_loss did not improve from 0.63028
Epoch 13/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6623 -
my_auc: 0.6425 - val_loss: 0.6850 - val_my_auc: 0.6370

Epoch 00013: val_loss did not improve from 0.63028
Epoch 14/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6616 -
my_auc: 0.6441 - val_loss: 0.6689 - val_my_auc: 0.6379

Epoch 00014: val_loss did not improve from 0.63028
Epoch 15/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6596 -
my_auc: 0.6508 - val_loss: 0.6990 - val_my_auc: 0.6379

Epoch 00015: val_loss did not improve from 0.63028
Epoch 16/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6594 -
my_auc: 0.6472 - val_loss: 0.6706 - val_my_auc: 0.6397
```

```
Epoch 00016: val_loss did not improve from 0.63028
Epoch 17/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6567 -
my_auc: 0.6566 - val_loss: 0.6701 - val_my_auc: 0.6406

Epoch 00017: val_loss did not improve from 0.63028
Epoch 18/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6557 -
my_auc: 0.6590 - val_loss: 0.7012 - val_my_auc: 0.6390

Epoch 00018: val_loss did not improve from 0.63028
Epoch 19/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6547 -
my_auc: 0.6596 - val_loss: 0.6702 - val_my_auc: 0.6417

Epoch 00019: val_loss did not improve from 0.63028
Epoch 20/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6516 -
my_auc: 0.6674 - val_loss: 0.6759 - val_my_auc: 0.6424

Epoch 00020: val_loss did not improve from 0.63028
Epoch 21/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6513 -
my_auc: 0.6673 - val_loss: 0.6975 - val_my_auc: 0.6419

Epoch 00021: val_loss did not improve from 0.63028
Epoch 22/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6499 -
my_auc: 0.6679 - val_loss: 0.6739 - val_my_auc: 0.6435

Epoch 00022: val_loss did not improve from 0.63028
Epoch 23/500
79/79 [==============================] - 3s 38ms/step - loss: 0.6488 -
my_auc: 0.6728 - val_loss: 0.7141 - val_my_auc: 0.6439

Epoch 00023: val_loss did not improve from 0.63028
Epoch 24/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6485 -
my_auc: 0.6731 - val_loss: 0.6928 - val_my_auc: 0.6444

Epoch 00024: val_loss did not improve from 0.63028
```

```
Epoch 25/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6468 -
my_auc: 0.6739 - val_loss: 0.6467 - val_my_auc: 0.6468

Epoch 00025: val_loss did not improve from 0.63028
Epoch 26/500
79/79 [==============================] - 3s 38ms/step - loss: 0.6433 -
my_auc: 0.6860 - val_loss: 0.6307 - val_my_auc: 0.6489

Epoch 00026: val_loss did not improve from 0.63028
Epoch 27/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6432 -
my_auc: 0.6866 - val_loss: 0.6604 - val_my_auc: 0.6489

Epoch 00027: val_loss did not improve from 0.63028
Epoch 28/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6407 -
my_auc: 0.6899 - val_loss: 0.6478 - val_my_auc: 0.6488

Epoch 00028: val_loss did not improve from 0.63028
Epoch 29/500
79/79 [==============================] - 3s 38ms/step - loss: 0.6385 -
my_auc: 0.6938 - val_loss: 0.7263 - val_my_auc: 0.6483

Epoch 00029: val_loss did not improve from 0.63028
Epoch 30/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6372 -
my_auc: 0.6992 - val_loss: 0.6897 - val_my_auc: 0.6502

Epoch 00030: val_loss did not improve from 0.63028
Epoch 31/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6354 -
my_auc: 0.6999 - val_loss: 0.6253 - val_my_auc: 0.6530

Epoch 00031: val_loss improved from 0.63028 to 0.62531, saving model t
o best_model_one.h5
Epoch 32/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6340 -
my_auc: 0.7062 - val_loss: 0.6922 - val_my_auc: 0.6525

Epoch 00032: val_loss did not improve from 0.62531
Epoch 33/500
```

```
79/79 [==============================] - 3s 36ms/step - loss: 0.6314 -
my_auc: 0.7094 - val_loss: 0.6377 - val_my_auc: 0.6550

Epoch 00033: val_loss did not improve from 0.62531
Epoch 34/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6298 -
my_auc: 0.7154 - val_loss: 0.6782 - val_my_auc: 0.6546

Epoch 00034: val_loss did not improve from 0.62531
Epoch 35/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6276 -
my_auc: 0.7150 - val_loss: 0.6661 - val_my_auc: 0.6563

Epoch 00035: val_loss did not improve from 0.62531
Epoch 36/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6250 -
my_auc: 0.7239 - val_loss: 0.6402 - val_my_auc: 0.6577

Epoch 00036: val_loss did not improve from 0.62531
Epoch 37/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6216 -
my_auc: 0.7329 - val_loss: 0.6652 - val_my_auc: 0.6578

Epoch 00037: val_loss did not improve from 0.62531
Epoch 38/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6199 -
my_auc: 0.7337 - val_loss: 0.6921 - val_my_auc: 0.6579

Epoch 00038: val_loss did not improve from 0.62531
Epoch 39/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6181 -
my_auc: 0.7343 - val_loss: 0.6671 - val_my_auc: 0.6596

Epoch 00039: val_loss did not improve from 0.62531
Epoch 40/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6152 -
my_auc: 0.7409 - val_loss: 0.6753 - val_my_auc: 0.6606

Epoch 00040: val_loss did not improve from 0.62531
Epoch 41/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6115 -
my_auc: 0.7475 - val_loss: 0.6919 - val_my_auc: 0.6605
```

```
Epoch 00041: val_loss did not improve from 0.62531
Epoch 42/500
79/79 [==============================] - 3s 37ms/step - loss: 0.6082 -
my_auc: 0.7555 - val_loss: 0.6177 - val_my_auc: 0.6638

Epoch 00042: val_loss improved from 0.62531 to 0.61771, saving model t
o best_model_one.h5
Epoch 43/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6078 -
my_auc: 0.7536 - val_loss: 0.6891 - val_my_auc: 0.6634

Epoch 00043: val_loss did not improve from 0.61771
Epoch 44/500
79/79 [==============================] - 3s 36ms/step - loss: 0.6037 -
my_auc: 0.7602 - val_loss: 0.6559 - val_my_auc: 0.6649

Epoch 00044: val_loss did not improve from 0.61771
Epoch 45/500
79/79 [==============================] - 3s 38ms/step - loss: 0.6011 -
my_auc: 0.7669 - val_loss: 0.6371 - val_my_auc: 0.6654

Epoch 00045: val_loss did not improve from 0.61771
Epoch 46/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5962 -
my_auc: 0.7729 - val_loss: 0.7066 - val_my_auc: 0.6647

Epoch 00046: val_loss did not improve from 0.61771
Epoch 47/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5957 -
my_auc: 0.7721 - val_loss: 0.6669 - val_my_auc: 0.6663

Epoch 00047: val_loss did not improve from 0.61771
Epoch 48/500
79/79 [==============================] - 3s 38ms/step - loss: 0.5917 -
my_auc: 0.7778 - val_loss: 0.6293 - val_my_auc: 0.6680

Epoch 00048: val_loss did not improve from 0.61771
Epoch 49/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5891 -
my_auc: 0.7819 - val_loss: 0.6552 - val_my_auc: 0.6681
```

```
Epoch 00049: val_loss did not improve from 0.61771
Epoch 50/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5856 -
my_auc: 0.7897 - val_loss: 0.6415 - val_my_auc: 0.6696

Epoch 00050: val_loss did not improve from 0.61771
Epoch 51/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5835 -
my_auc: 0.7905 - val_loss: 0.6705 - val_my_auc: 0.6698

Epoch 00051: val_loss did not improve from 0.61771
Epoch 52/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5794 -
my_auc: 0.7967 - val_loss: 0.6552 - val_my_auc: 0.6708

Epoch 00052: val_loss did not improve from 0.61771
Epoch 53/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5763 -
my_auc: 0.8019 - val_loss: 0.6784 - val_my_auc: 0.6720

Epoch 00053: val_loss did not improve from 0.61771
Epoch 54/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5718 -
my_auc: 0.8075 - val_loss: 0.6858 - val_my_auc: 0.6722

Epoch 00054: val_loss did not improve from 0.61771
Epoch 55/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5716 -
my_auc: 0.8062 - val_loss: 0.6154 - val_my_auc: 0.6743

Epoch 00055: val_loss improved from 0.61771 to 0.61538, saving model t
o best_model_one.h5
Epoch 56/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5661 -
my_auc: 0.8139 - val_loss: 0.6571 - val_my_auc: 0.6745

Epoch 00056: val_loss did not improve from 0.61538
Epoch 57/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5631 -
my_auc: 0.8202 - val_loss: 0.7061 - val_my_auc: 0.6742

Epoch 00057: val_loss did not improve from 0.61538
```

```
Epoch 58/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5609 -
my_auc: 0.8189 - val_loss: 0.6714 - val_my_auc: 0.6756

Epoch 00058: val_loss did not improve from 0.61538
Epoch 59/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5556 -
my_auc: 0.8268 - val_loss: 0.6996 - val_my_auc: 0.6759

Epoch 00059: val_loss did not improve from 0.61538
Epoch 60/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5526 -
my_auc: 0.8281 - val_loss: 0.6553 - val_my_auc: 0.6767

Epoch 00060: val_loss did not improve from 0.61538
Epoch 61/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5493 -
my_auc: 0.8342 - val_loss: 0.6852 - val_my_auc: 0.6776

Epoch 00061: val_loss did not improve from 0.61538
Epoch 62/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5439 -
my_auc: 0.8412 - val_loss: 0.6625 - val_my_auc: 0.6782

Epoch 00062: val_loss did not improve from 0.61538
Epoch 63/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5418 -
my_auc: 0.8387 - val_loss: 0.6803 - val_my_auc: 0.6792

Epoch 00063: val_loss did not improve from 0.61538
Epoch 64/500
79/79 [==============================] - 3s 38ms/step - loss: 0.5389 -
my_auc: 0.8447 - val_loss: 0.6507 - val_my_auc: 0.6798

Epoch 00064: val_loss did not improve from 0.61538
Epoch 65/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5344 -
my_auc: 0.8484 - val_loss: 0.6687 - val_my_auc: 0.6801

Epoch 00065: val_loss did not improve from 0.61538
Epoch 66/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5308 -
```

```
my_auc: 0.8524 - val_loss: 0.6524 - val_my_auc: 0.6813

Epoch 00066: val_loss did not improve from 0.61538
Epoch 67/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5271 -
my_auc: 0.8546 - val_loss: 0.6179 - val_my_auc: 0.6821

Epoch 00067: val_loss did not improve from 0.61538
Epoch 68/500
79/79 [==============================] - 3s 38ms/step - loss: 0.5248 -
my_auc: 0.8590 - val_loss: 0.6605 - val_my_auc: 0.6823

Epoch 00068: val_loss did not improve from 0.61538
Epoch 69/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5186 -
my_auc: 0.8616 - val_loss: 0.7047 - val_my_auc: 0.6825

Epoch 00069: val_loss did not improve from 0.61538
Epoch 70/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5157 -
my_auc: 0.8695 - val_loss: 0.6233 - val_my_auc: 0.6843

Epoch 00070: val_loss did not improve from 0.61538
Epoch 71/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5116 -
my_auc: 0.8689 - val_loss: 0.6604 - val_my_auc: 0.6840

Epoch 00071: val_loss did not improve from 0.61538
Epoch 72/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5086 -
my_auc: 0.8705 - val_loss: 0.6236 - val_my_auc: 0.6849

Epoch 00072: val_loss did not improve from 0.61538
Epoch 73/500
79/79 [==============================] - 3s 37ms/step - loss: 0.5057 -
my_auc: 0.8747 - val_loss: 0.6798 - val_my_auc: 0.6852

Epoch 00073: val_loss did not improve from 0.61538
Epoch 74/500
79/79 [==============================] - 3s 36ms/step - loss: 0.5022 -
my_auc: 0.8750 - val_loss: 0.6367 - val_my_auc: 0.6856
```

```
Epoch 00074: val_loss did not improve from 0.61538
Epoch 75/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4963 -
my_auc: 0.8807 - val_loss: 0.6070 - val_my_auc: 0.6856

Epoch 00075: val_loss improved from 0.61538 to 0.60704, saving model t
o best_model_one.h5
Epoch 76/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4925 -
my_auc: 0.8849 - val_loss: 0.6416 - val_my_auc: 0.6864

Epoch 00076: val_loss did not improve from 0.60704
Epoch 77/500
79/79 [==============================] - 3s 37ms/step - loss: 0.4884 -
my_auc: 0.8877 - val_loss: 0.6661 - val_my_auc: 0.6865

Epoch 00077: val_loss did not improve from 0.60704
Epoch 78/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4856 -
my_auc: 0.8890 - val_loss: 0.6578 - val_my_auc: 0.6873

Epoch 00078: val_loss did not improve from 0.60704
Epoch 79/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4823 -
my_auc: 0.8925 - val_loss: 0.6971 - val_my_auc: 0.6871

Epoch 00079: val_loss did not improve from 0.60704
Epoch 80/500
79/79 [==============================] - 3s 39ms/step - loss: 0.4771 -
my_auc: 0.8931 - val_loss: 0.6956 - val_my_auc: 0.6868

Epoch 00080: val_loss did not improve from 0.60704
Epoch 81/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4737 -
my_auc: 0.8973 - val_loss: 0.6444 - val_my_auc: 0.6876

Epoch 00081: val_loss did not improve from 0.60704
Epoch 82/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4716 -
my_auc: 0.9010 - val_loss: 0.6312 - val_my_auc: 0.6887

Epoch 00082: val_loss did not improve from 0.60704
```

```
Epoch 83/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4644 -
my_auc: 0.9047 - val_loss: 0.6306 - val_my_auc: 0.6893

Epoch 00083: val_loss did not improve from 0.60704
Epoch 84/500
79/79 [==============================] - 3s 38ms/step - loss: 0.4625 -
my_auc: 0.9072 - val_loss: 0.6102 - val_my_auc: 0.6901

Epoch 00084: val_loss did not improve from 0.60704
Epoch 85/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4573 -
my_auc: 0.9096 - val_loss: 0.6315 - val_my_auc: 0.6897

Epoch 00085: val_loss did not improve from 0.60704
Epoch 86/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4522 -
my_auc: 0.9141 - val_loss: 0.5979 - val_my_auc: 0.6910

Epoch 00086: val_loss improved from 0.60704 to 0.59787, saving model t
o best_model_one.h5
Epoch 87/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4476 -
my_auc: 0.9157 - val_loss: 0.6613 - val_my_auc: 0.6903

Epoch 00087: val_loss did not improve from 0.59787
Epoch 88/500
79/79 [==============================] - 3s 37ms/step - loss: 0.4432 -
my_auc: 0.9186 - val_loss: 0.6363 - val_my_auc: 0.6911

Epoch 00088: val_loss did not improve from 0.59787
Epoch 89/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4403 -
my_auc: 0.9219 - val_loss: 0.6144 - val_my_auc: 0.6912

Epoch 00089: val_loss did not improve from 0.59787
Epoch 90/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4359 -
my_auc: 0.9234 - val_loss: 0.6348 - val_my_auc: 0.6914

Epoch 00090: val_loss did not improve from 0.59787
Epoch 91/500
```

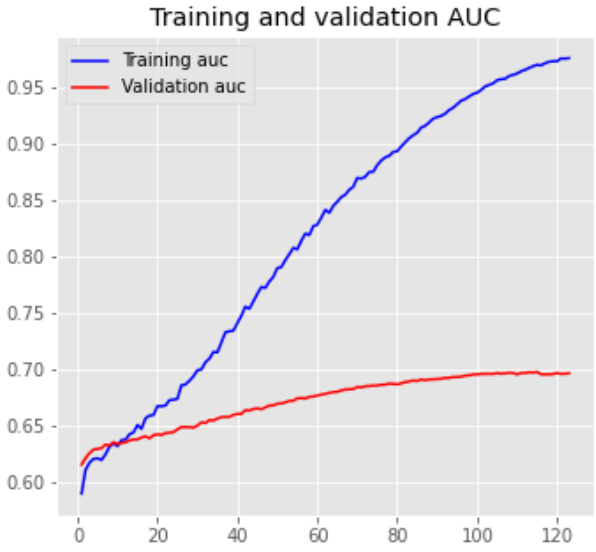
```
79/79 [==============================] - 3s 37ms/step - loss: 0.4324 -
my_auc: 0.9243 - val_loss: 0.6683 - val_my_auc: 0.6917

Epoch 00091: val_loss did not improve from 0.59787
Epoch 92/500
79/79 [==============================] - 3s 37ms/step - loss: 0.4292 -
my_auc: 0.9263 - val_loss: 0.6511 - val_my_auc: 0.6924

Epoch 00092: val_loss did not improve from 0.59787
Epoch 93/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4242 -
my_auc: 0.9296 - val_loss: 0.6642 - val_my_auc: 0.6925

Epoch 00093: val_loss did not improve from 0.59787
Epoch 94/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4188 -
my_auc: 0.9316 - val_loss: 0.6411 - val_my_auc: 0.6932

Epoch 00094: val_loss did not improve from 0.59787
Epoch 95/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4170 -
my_auc: 0.9347 - val_loss: 0.6501 - val_my_auc: 0.6931

Epoch 00095: val_loss did not improve from 0.59787
Epoch 96/500
79/79 [==============================] - 3s 37ms/step - loss: 0.4116 -
my_auc: 0.9379 - val_loss: 0.6609 - val_my_auc: 0.6935

Epoch 00096: val_loss did not improve from 0.59787
Epoch 97/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4067 -
my_auc: 0.9393 - val_loss: 0.6281 - val_my_auc: 0.6943

Epoch 00097: val_loss did not improve from 0.59787
Epoch 98/500
79/79 [==============================] - 3s 36ms/step - loss: 0.4033 -
my_auc: 0.9416 - val_loss: 0.5943 - val_my_auc: 0.6946

Epoch 00098: val_loss improved from 0.59787 to 0.59434, saving model t
o best_model_one.h5
Epoch 99/500
79/79 [==============================] - 3s 37ms/step - loss: 0.3991 -
```

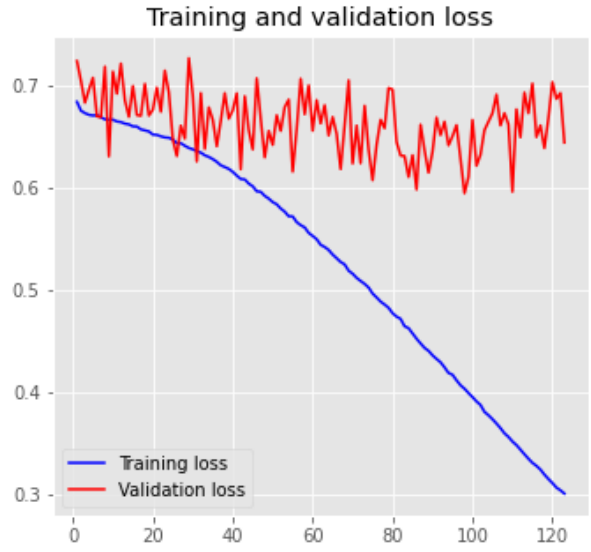
```
my_auc: 0.9438 - val_loss: 0.6103 - val_my_auc: 0.6955

Epoch 00099: val_loss did not improve from 0.59434
Epoch 100/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3951 -
my_auc: 0.9450 - val_loss: 0.6660 - val_my_auc: 0.6955

Epoch 00100: val_loss did not improve from 0.59434
Epoch 101/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3909 -
my_auc: 0.9473 - val_loss: 0.6211 - val_my_auc: 0.6959

Epoch 00101: val_loss did not improve from 0.59434
Epoch 102/500
79/79 [==============================] - 3s 38ms/step - loss: 0.3875 -
my_auc: 0.9505 - val_loss: 0.6320 - val_my_auc: 0.6960

Epoch 00102: val_loss did not improve from 0.59434
Epoch 103/500
79/79 [==============================] - 3s 37ms/step - loss: 0.3807 -
my_auc: 0.9518 - val_loss: 0.6559 - val_my_auc: 0.6960

Epoch 00103: val_loss did not improve from 0.59434
Epoch 104/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3774 -
my_auc: 0.9532 - val_loss: 0.6643 - val_my_auc: 0.6960

Epoch 00104: val_loss did not improve from 0.59434
Epoch 105/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3735 -
my_auc: 0.9560 - val_loss: 0.6725 - val_my_auc: 0.6966

Epoch 00105: val_loss did not improve from 0.59434
Epoch 106/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3694 -
my_auc: 0.9569 - val_loss: 0.6907 - val_my_auc: 0.6961

Epoch 00106: val_loss did not improve from 0.59434
Epoch 107/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3649 -
my_auc: 0.9572 - val_loss: 0.6606 - val_my_auc: 0.6964
```

```
Epoch 00107: val_loss did not improve from 0.59434
Epoch 108/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3598 -
my_auc: 0.9597 - val_loss: 0.6725 - val_my_auc: 0.6969

Epoch 00108: val_loss did not improve from 0.59434
Epoch 109/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3563 -
my_auc: 0.9608 - val_loss: 0.6617 - val_my_auc: 0.6968

Epoch 00109: val_loss did not improve from 0.59434
Epoch 110/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3518 -
my_auc: 0.9619 - val_loss: 0.5958 - val_my_auc: 0.6955

Epoch 00110: val_loss did not improve from 0.59434
Epoch 111/500
79/79 [==============================] - 3s 37ms/step - loss: 0.3483 -
my_auc: 0.9640 - val_loss: 0.6763 - val_my_auc: 0.6968

Epoch 00111: val_loss did not improve from 0.59434
Epoch 112/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3438 -
my_auc: 0.9651 - val_loss: 0.6490 - val_my_auc: 0.6967

Epoch 00112: val_loss did not improve from 0.59434
Epoch 113/500
79/79 [==============================] - 3s 38ms/step - loss: 0.3393 -
my_auc: 0.9669 - val_loss: 0.6926 - val_my_auc: 0.6974

Epoch 00113: val_loss did not improve from 0.59434
Epoch 114/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3350 -
my_auc: 0.9684 - val_loss: 0.6725 - val_my_auc: 0.6971

Epoch 00114: val_loss did not improve from 0.59434
Epoch 115/500
79/79 [==============================] - 3s 37ms/step - loss: 0.3309 -
my_auc: 0.9696 - val_loss: 0.7015 - val_my_auc: 0.6977

Epoch 00115: val_loss did not improve from 0.59434
Epoch 116/500
```

```
79/79 [==============================] - 3s 36ms/step - loss: 0.3280 -
my_auc: 0.9692 - val_loss: 0.6492 - val_my_auc: 0.6955

Epoch 00116: val_loss did not improve from 0.59434
Epoch 117/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3244 -
my_auc: 0.9710 - val_loss: 0.6606 - val_my_auc: 0.6956

Epoch 00117: val_loss did not improve from 0.59434
Epoch 118/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3195 -
my_auc: 0.9723 - val_loss: 0.6384 - val_my_auc: 0.6955

Epoch 00118: val_loss did not improve from 0.59434
Epoch 119/500
79/79 [==============================] - 3s 37ms/step - loss: 0.3152 -
my_auc: 0.9728 - val_loss: 0.6675 - val_my_auc: 0.6958

Epoch 00119: val_loss did not improve from 0.59434
Epoch 120/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3112 -
my_auc: 0.9727 - val_loss: 0.7028 - val_my_auc: 0.6967

Epoch 00120: val_loss did not improve from 0.59434
Epoch 121/500
79/79 [==============================] - 3s 38ms/step - loss: 0.3068 -
my_auc: 0.9753 - val_loss: 0.6864 - val_my_auc: 0.6959

Epoch 00121: val_loss did not improve from 0.59434
Epoch 122/500
79/79 [==============================] - 3s 38ms/step - loss: 0.3041 -
my_auc: 0.9752 - val_loss: 0.6922 - val_my_auc: 0.6961

Epoch 00122: val_loss did not improve from 0.59434
Epoch 123/500
79/79 [==============================] - 3s 36ms/step - loss: 0.3009 -
my_auc: 0.9756 - val_loss: 0.6438 - val_my_auc: 0.6964

Epoch 00123: val_loss did not improve from 0.59434
Epoch 00123: early stopping
166/166 [==============================] - 1s 4ms/step - loss: 0.6050
```

```
- my_auc: 0.6940
AUC: 0.6939592957496643
```

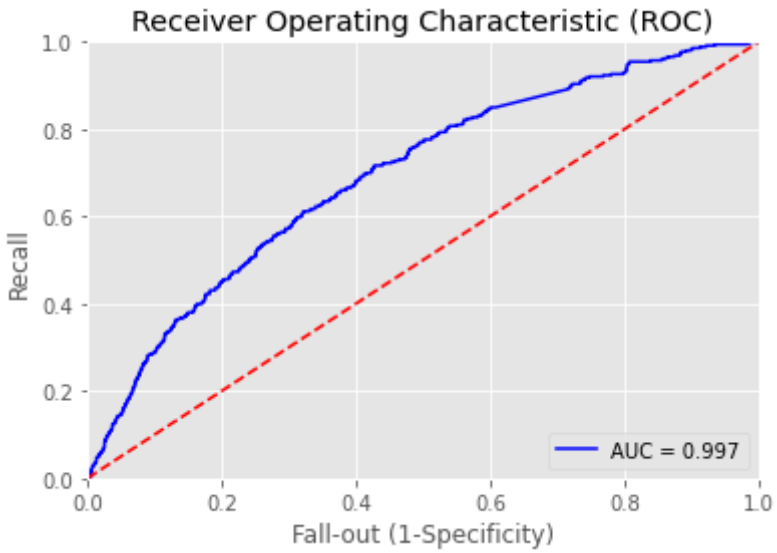


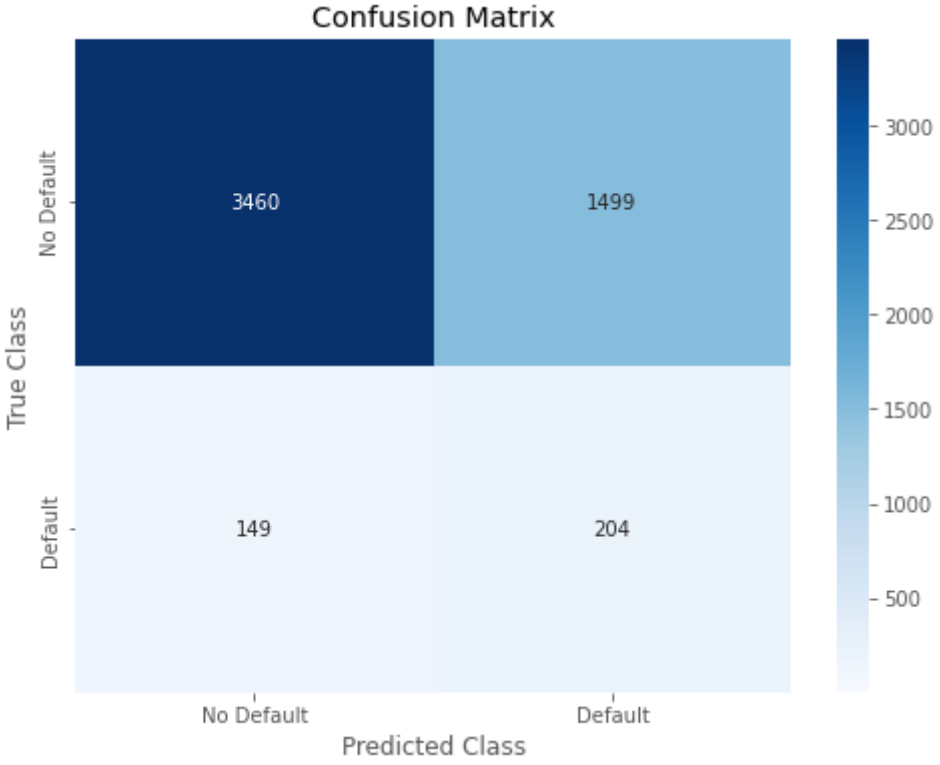


In [51]:

```
#predictions, confusion matrix, and ROC curve for the second model
predictions_NN_prob = saved_model_one.predict(X_test_sequences)
predictions_NN_prob = predictions_NN_prob[:,0]
predictions_NN_01 = np.where (predictions_NN_prob >.5,1,0)
confusion_matrix(y_test, predictions_NN_01)
print(accuracy_score(y_test, predictions_NN_01, normalize=False) / float(
y_test.size))
false_positive_rate, recall, thresholds = roc_curve(y_test, predictions_N
N_prob)
roc_auc = auc(false_positive_rate, recall)
plt.figure()
plt.title('Receiver Operating Characteristic (ROC)')
plt.plot(false_positive_rate, recall, 'b', label = 'AUC = %0.3f' %roc_auc
)
plt.legend(loc='lower right')
plt.plot([0,1], [0,1], 'r--')
plt.xlim([0.0,1.0])
plt.ylim([0.0,1.0])
plt.ylabel('Recall')
plt.xlabel('Fall-out (1-Specificity)')
plt.show()
cm = confusion_matrix(y_test, predictions_NN_01)
labels = ['No Default', 'Default']
plt.figure(figsize=(8,6))
sns.heatmap(cm,xticklabels=labels, yticklabels=labels, annot=True, fmt=
'd', cmap="Blues", vmin = 0.2);
plt.title('Confusion Matrix')
plt.ylabel('True Class')
plt.xlabel('Predicted Class')
plt.show()
```

```
0.6897590361445783
```





In [52]:

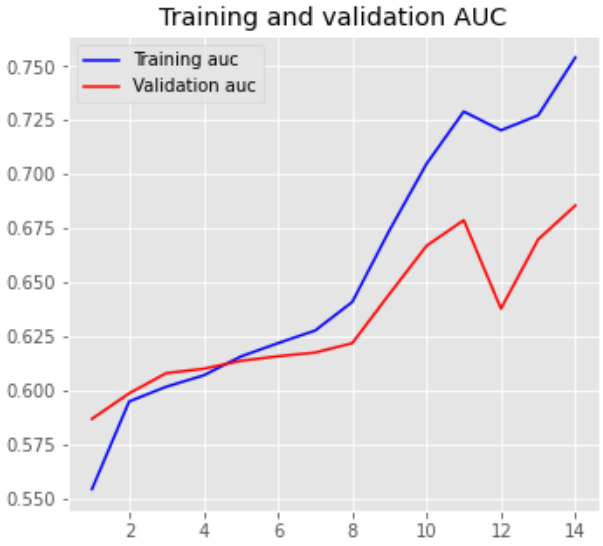
```
#LSTM model
with tf.device('/device:GPU:0'):
    model = Sequential()
    model.add(Embedding(len(vect.get_feature_names()) + 1,
                    128,
                    input_length=MAX_SEQ_LENGTH))
    model.add(LSTM(128))
    model.add(Dropout(0.2))
    model.add(Dense(50, activation='relu'))
    model.add(Dense(units=1, activation='sigmoid'))
    Adam = tf.keras.optimizers.Adam(lr=0.00003)
    model.compile(loss='binary_crossentropy', optimizer= Adam, metrics=[t
f.keras.metrics.AUC(name='my_auc')])
    print(model.summary())
    es = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patienc
e=5)
    mc = ModelCheckpoint('best_model_two.h5', monitor='val_loss', mode='m
in', verbose=1, save_best_only=True)
    hi = model.fit(X_train_sequences, y_train,
          epochs=50, batch_size=64, verbose=1,
          validation_data=(X_valid_sequences, y_valid),callbacks=[es,mc])
    saved_model_two = load_model('best_model_two.h5')
    scores = saved_model_two.evaluate(X_test_sequences, y_test, verbose=1
)
    print("AUC:", scores[1])

    plot_history(hi)
```

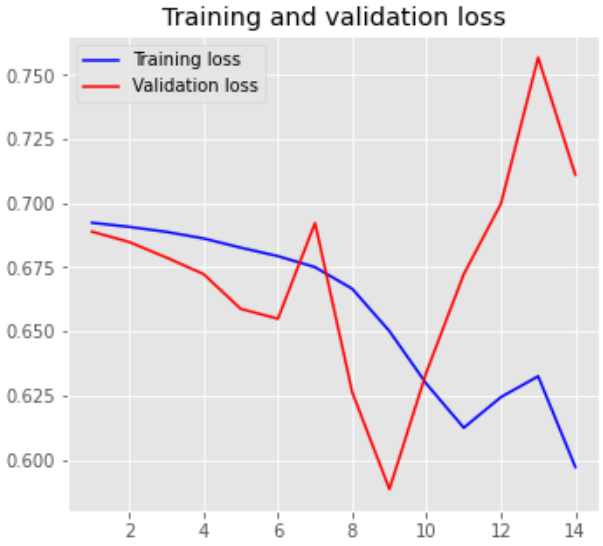
```
/opt/conda/lib/python3.7/site-packages/keras/optimizer_v2/optimizer_v
2.py:356: UserWarning: The `lr` argument is deprecated, use `learning_
rate` instead.
  "The `lr` argument is deprecated, use `learning_rate` instead.")
```

```
Model: "sequential_14"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_4 (Embedding)      (None, 4832, 128)         1619072
_________________________________________________________________
lstm (LSTM)                  (None, 128)               131584
_________________________________________________________________
dropout_44 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_58 (Dense)             (None, 50)                6450
_________________________________________________________________
dense_59 (Dense)             (None, 1)                 51
=================================================================
Total params: 1,757,157
Trainable params: 1,757,157
Non-trainable params: 0
_________________________________________________________________
None
Epoch 1/50
79/79 [==============================] - 30s 347ms/step - loss: 0.6924
- my_auc: 0.5542 - val_loss: 0.6889 - val_my_auc: 0.5867

Epoch 00001: val_loss improved from inf to 0.68889, saving model to be
st_model_two.h5
Epoch 2/50
79/79 [==============================] - 27s 342ms/step - loss: 0.6908
- my_auc: 0.5948 - val_loss: 0.6848 - val_my_auc: 0.5986

Epoch 00002: val_loss improved from 0.68889 to 0.68485, saving model t
o best_model_two.h5
Epoch 3/50
79/79 [==============================] - 27s 340ms/step - loss: 0.6888
- my_auc: 0.6016 - val_loss: 0.6788 - val_my_auc: 0.6079

Epoch 00003: val_loss improved from 0.68485 to 0.67881, saving model t
o best_model_two.h5
Epoch 4/50
79/79 [==============================] - 26s 337ms/step - loss: 0.6862
- my_auc: 0.6069 - val_loss: 0.6724 - val_my_auc: 0.6099
```

```
Epoch 00004: val_loss improved from 0.67881 to 0.67236, saving model t
o best_model_two.h5
Epoch 5/50
79/79 [==============================] - 27s 340ms/step - loss: 0.6826
- my_auc: 0.6156 - val_loss: 0.6588 - val_my_auc: 0.6136

Epoch 00005: val_loss improved from 0.67236 to 0.65879, saving model t
o best_model_two.h5
Epoch 6/50
79/79 [==============================] - 27s 339ms/step - loss: 0.6793
- my_auc: 0.6218 - val_loss: 0.6549 - val_my_auc: 0.6157

Epoch 00006: val_loss improved from 0.65879 to 0.65492, saving model t
o best_model_two.h5
Epoch 7/50
79/79 [==============================] - 30s 375ms/step - loss: 0.6750
- my_auc: 0.6276 - val_loss: 0.6923 - val_my_auc: 0.6174

Epoch 00007: val_loss did not improve from 0.65492
Epoch 8/50
79/79 [==============================] - 27s 342ms/step - loss: 0.6666
- my_auc: 0.6408 - val_loss: 0.6263 - val_my_auc: 0.6218

Epoch 00008: val_loss improved from 0.65492 to 0.62626, saving model t
o best_model_two.h5
Epoch 9/50
79/79 [==============================] - 27s 343ms/step - loss: 0.6501
- my_auc: 0.6740 - val_loss: 0.5885 - val_my_auc: 0.6445

Epoch 00009: val_loss improved from 0.62626 to 0.58851, saving model t
o best_model_two.h5
Epoch 10/50
79/79 [==============================] - 27s 339ms/step - loss: 0.6295
- my_auc: 0.7049 - val_loss: 0.6343 - val_my_auc: 0.6669

Epoch 00010: val_loss did not improve from 0.58851
Epoch 11/50
79/79 [==============================] - 27s 344ms/step - loss: 0.6124
- my_auc: 0.7290 - val_loss: 0.6723 - val_my_auc: 0.6788

Epoch 00011: val_loss did not improve from 0.58851
Epoch 12/50
```

```
79/79 [==============================] - 27s 341ms/step - loss: 0.6244
- my_auc: 0.7204 - val_loss: 0.6998 - val_my_auc: 0.6378

Epoch 00012: val_loss did not improve from 0.58851
Epoch 13/50
79/79 [==============================] - 27s 342ms/step - loss: 0.6325
- my_auc: 0.7273 - val_loss: 0.7568 - val_my_auc: 0.6698

Epoch 00013: val_loss did not improve from 0.58851
Epoch 14/50
79/79 [==============================] - 27s 339ms/step - loss: 0.5971
- my_auc: 0.7541 - val_loss: 0.7110 - val_my_auc: 0.6855

Epoch 00014: val_loss did not improve from 0.58851
Epoch 00014: early stopping
166/166 [==============================] - 13s 76ms/step - loss: 0.589
1 - my_auc: 0.6602
AUC: 0.6602194309234619
```





In [53]:

```
#predictions, confusion matrix, and ROC curve for the third model
predictions_NN_prob = saved_model_two.predict(X_test_sequences)
predictions_NN_prob = predictions_NN_prob[:,0]
predictions_NN_01 = np.where (predictions_NN_prob >.5,1,0)
confusion_matrix(y_test, predictions_NN_01)
print(accuracy_score(y_test, predictions_NN_01, normalize=False) / float(
y_test.size))
false_positive_rate, recall, thresholds = roc_curve(y_test, predictions_N
N_prob)
roc_auc = auc(false_positive_rate, recall)
plt.figure()
plt.title('Receiver Operating Characteristic (ROC)')
plt.plot(false_positive_rate, recall, 'b', label = 'AUC = %0.3f' %roc_auc
)
plt.legend(loc='lower right')
plt.plot([0,1], [0,1], 'r--')
plt.xlim([0.0,1.0])
plt.ylim([0.0,1.0])
plt.ylabel('Recall')
plt.xlabel('Fall-out (1-Specificity)')
plt.show()
cm = confusion_matrix(y_test, predictions_NN_01)
labels = ['No Default', 'Default']
plt.figure(figsize=(8,6))
sns.heatmap(cm,xticklabels=labels, yticklabels=labels, annot=True, fmt=
'd', cmap="Blues", vmin = 0.2);
plt.title('Confusion Matrix')
plt.ylabel('True Class')
plt.xlabel('Predicted Class')
plt.show()
```

```
0.7447289156626506
```