```
In [1]: import pandas as pd
        import numpy as np
        import tensorflow as tf
        import matplotlib.pyplot as plt
        import os
        import random
        import cv2
        import imageio
        import sklearn
        from sklearn.model selection import train test split
        from PIL import Image
        from os import listdir
        from tqdm import tqdm notebook
        from skimage.color import rgb2gray
        from collections import Counter
        from sklearn.metrics import roc_auc_score,accuracy_score
        from tensorflow.keras import backend as K
        from tensorflow.keras.layers import Conv2D, MaxPooling2D, Dropout, Flatten, Dense
        from tensorflow.keras.models import Sequential
        from tensorflow.keras.preprocessing.image import ImageDataGenerator
        random_seed = 33
```

Load Images and Resize them

```
In [2]: # cell labels are 1 if infected, 0 if uninfected.
        cell labels = []
        # cells image
        cell images = []
        def read_images(file_path, label):
            images addresses=os.listdir(file path)
            images addresses.remove('Thumbs.db')
            cells image array = np.zeros((len(images addresses),125,125,3),dtype=np.int16)
            for ind, img address in tqdm notebook(enumerate(images addresses), total=len(imag
        es addresses)):
                path = file_path +'/'+ img_address
                image = Image.open(path)
                image = image.resize((125, 125))
                image_array = np.asarray(image)
                image_array = image_array.astype(np.int32)
                cells image array[ind]=image array
                cell images.append(image array)
                cell_labels.append(label)
            return cells_image_array
```

```
In [3]: infected_cells = read_images("D:/Neural/Project/cell_images/cell_images/Parasitize
d", 1)
    uninfected_cells = read_images("D:/Neural/Project/cell_images/cell_images/Uninfecte
d",0)

D:\Anaconda\envs\TensorFlow-GPU\lib\site-packages\ipykernel_launcher.py:12: Tqdm
    DeprecationWarning: This function will be removed in tqdm==5.0.0
    Please use `tqdm.notebook.tqdm` instead of `tqdm.tqdm_notebook`
    if sys.path[0] == '':
```

Sample Images of Infected and Uninfected Cells

```
In [4]: def SampleCells(images_arr,title):
    fig, axes = plt.subplots(1, 10, figsize=(20,20))

for img, ax in zip( images_arr, axes):
    ax.imshow(img)
    ax.set_title(title,fontsize=20)
    plt.tight_layout()
    plt.show()
In [5]: SampleCells(infected_cells,'Infected')
SampleCells(uninfected_cells,'Uninfected')

Uninfected

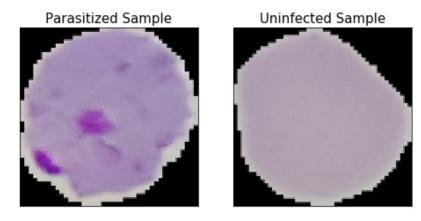
Uninfected
```

Data Exploration

select one sample from each category - Infected and Uninfected Cells

```
In [6]: np.random.seed(random_seed)
        rand=np.random.randint(infected cells.shape[0])
        parasitized_sample=infected_cells[rand]
        np.random.seed(random_seed)
        rand=np.random.randint(uninfected_cells.shape[0])
        uninfected sample=uninfected cells[rand]
        fig=plt.figure(figsize=(8,5))
        plt.title("Sample Images for Reference", fontsize=20)
        plt.axis('off')
        ax1=fig.add subplot(121)
        ax1.imshow(parasitized sample)
        ax1.get_xaxis().set_visible(False)
        ax1.get_yaxis().set_visible(False)
        ax1.set_title("Parasitized Sample", fontsize=15)
        ax2=fig.add_subplot(122)
        ax2.imshow(uninfected_sample)
        ax2.get xaxis().set visible(False)
        ax2.get yaxis().set visible(False)
        ax2.set_title("Uninfected Sample", fontsize=15)
        plt.show()
```

Sample Images for Reference

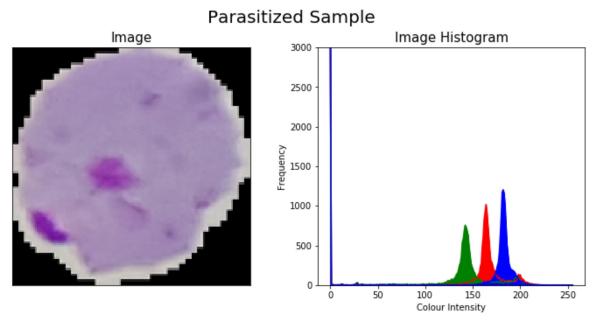


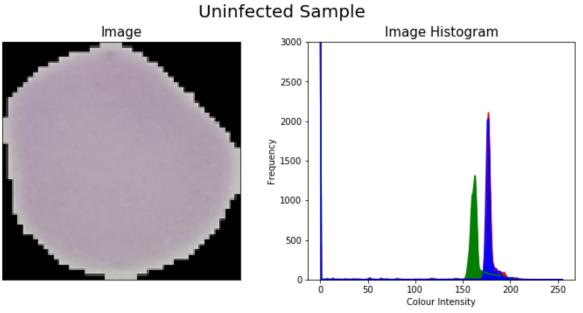
Analyse Blood Cell Color

To do so we plot the image histogram which shows Red-Green-Blue (RGB) color distribution for an image

```
In [7]: def get_RGB_hist(sample_data,title,to_return=False):
            sample=Image.fromarray(sample data.astype(np.uint8), 'RGB')
            hist=sample.histogram()
            hist_r=hist[:256]
            hist g=hist[256:512]
            hist b=hist[512:]
            if(to return==True):
                return hist_r,hist_g,hist_b
            x=np.arange(256)
            fig=plt.figure(figsize=(12,5))
            plt.title(title+'\n', fontsize=20)
            plt.axis('off')
            ax1=fig.add subplot(121)
            ax1.imshow(sample)
            ax1.set_title("Image", fontsize=15)
            ax1.get_xaxis().set_visible(False)
            ax1.get yaxis().set visible(False)
            ax2=fig.add subplot(122)
            ax2.plot(hist_r,color='red')
            ax2.fill_between(x,hist_r,color='red')
            ax2.plot(hist g,color='green')
            ax2.fill between(x,hist g,color='green')
            ax2.plot(hist_b,color='blue')
            ax2.fill_between(x,hist_b,color='blue')
            ax2.set_ylim(0,3000)
            ax2.set title("Image Histogram", fontsize=15)
            ax2.set ylabel("Frequency")
            ax2.set_xlabel("Colour Intensity")
            plt.show()
```

```
In [8]: get_RGB_hist(parasitized_sample, "Parasitized Sample")
get_RGB_hist(uninfected_sample, "Uninfected Sample")
```





Training, Testing and Validation Data Preparation

```
In [9]: def reordering_list(old_list,order):
    new_list = []
    for i in order:
        new_list.append(old_list[i])
    return new_list
```

```
In [10]: np.random.seed(seed=42)
         indices = np.arange(len(cell labels))
         np.random.shuffle(indices)
         indices = indices.tolist()
         cell labels = reordering list(cell labels,indices)
         cell_images = reordering_list(cell images,indices)
         #change to arrays
         image array = np.array(cell images)
         label array = np.array(cell labels)
In [11]: X_train, X_test, y_train, y_test = train_test_split(image_array, label_array, test_
         size=0.3, random state=100)
         X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.1,
         random state=100)
In [12]: print(X train.shape, X val.shape, X test.shape)
         print('Train:', Counter(y train), '\nVal:', Counter(y val), '\nTest:', Counter(y te
         st))
         (17361, 125, 125, 3) (1929, 125, 125, 3) (8268, 125, 125, 3)
         Train: Counter({1: 8703, 0: 8658})
         Val: Counter({1: 978, 0: 951})
         Test: Counter({0: 4170, 1: 4098})
```

CNN Basic Model

Label Encoding (One Hot Encoding)

```
In [13]: y_train=tf.keras.utils.to_categorical(y_train,2)
    y_test=tf.keras.utils.to_categorical(y_test,2)
    y_val=tf.keras.utils.to_categorical(y_val,2)
```

Model Hyperparameter Values

```
In [14]: EPOCHS = 20
BATCH_SIZE = 64
```

CNN Basic Model Architecture

```
In [15]: model cnn = Sequential()
         model_cnn.add(Conv2D(filters=16, kernel_size=2, padding="same", activation="relu", inpu
         t_shape=(125,125,3)))
         model_cnn.add(MaxPooling2D(pool_size=2))
         model_cnn.add(Conv2D(filters=32,kernel_size=2,padding="same",activation="relu"))
         model cnn.add(MaxPooling2D(pool size=2))
         model_cnn.add(Conv2D(filters=64,kernel_size=2,padding="same",activation="relu"))
         model cnn.add(MaxPooling2D(pool size=2))
         model cnn.add(Dropout(0.2))
         model cnn.add(Flatten())
         model cnn.add(Dense(500, activation="relu"))
         model cnn.add(Dropout(0.2))
         model cnn.add(Dense(2,activation="softmax")) #2 represent output layer neurons
In [16]: model cnn.compile(loss='categorical crossentropy', optimizer='adam', metrics=['accu
         racy'])
         model_cnn.summary()
         Model: "sequential"
         Layer (type)
                                     Output Shape
                                                              Param #
         conv2d (Conv2D)
                                     (None, 125, 125, 16)
                                                               208
         max pooling2d (MaxPooling2D) (None, 62, 62, 16)
                                                               2080
         conv2d 1 (Conv2D)
                                      (None, 62, 62, 32)
         max pooling2d 1 (MaxPooling2 (None, 31, 31, 32)
         conv2d 2 (Conv2D)
                                     (None, 31, 31, 64)
                                                               8256
         max_pooling2d_2 (MaxPooling2 (None, 15, 15, 64)
         dropout (Dropout)
                                     (None, 15, 15, 64)
                                     (None, 14400)
         flatten (Flatten)
         dense (Dense)
                                     (None, 500)
                                                               7200500
         dropout 1 (Dropout)
                                     (None, 500)
         dense_1 (Dense)
                                     (None, 2)
                                                               1002
         _____
         Total params: 7,212,046
         Trainable params: 7,212,046
         Non-trainable params: 0
```

Train the Model

```
Train on 17361 samples, validate on 1929 samples
Epoch 1/20
- accuracy: 0.5608 - val loss: 0.6507 - val accuracy: 0.6195
Epoch 2/20
accuracy: 0.6527 - val loss: 0.6003 - val accuracy: 0.6739
accuracy: 0.6920 - val loss: 0.5764 - val accuracy: 0.7045
Epoch 4/20
accuracy: 0.7278 - val loss: 0.5322 - val accuracy: 0.7429
Epoch 5/20
accuracy: 0.7932 - val loss: 0.4322 - val accuracy: 0.7880
accuracy: 0.8642 - val loss: 0.3577 - val accuracy: 0.8455
Epoch 7/20
accuracy: 0.8986 - val_loss: 0.2944 - val_accuracy: 0.8828
Epoch 8/20
accuracy: 0.9199 - val loss: 0.2447 - val accuracy: 0.9005
Epoch 9/20
17361/17361 [============== ] - 8s 471us/sample - loss: 0.1629 -
accuracy: 0.9355 - val loss: 0.2473 - val accuracy: 0.9129
Epoch 10/20
accuracy: 0.9434 - val loss: 0.2511 - val accuracy: 0.9145
Epoch 11/20
accuracy: 0.9497 - val loss: 0.2357 - val accuracy: 0.9222
Epoch 12/20
accuracy: 0.9545 - val loss: 0.2278 - val accuracy: 0.9155
Epoch 13/20
17361/17361 [============== ] - 8s 473us/sample - loss: 0.1006 -
accuracy: 0.9593 - val loss: 0.2255 - val accuracy: 0.9279
Epoch 14/20
accuracy: 0.9611 - val_loss: 0.2183 - val_accuracy: 0.9222
Epoch 15/20
accuracy: 0.9670 - val loss: 0.2562 - val accuracy: 0.9248
Epoch 16/20
17361/17361 [============== ] - 8s 473us/sample - loss: 0.0730 -
accuracy: 0.9700 - val loss: 0.2524 - val accuracy: 0.9342
Epoch 17/20
accuracy: 0.9696 - val_loss: 0.2358 - val_accuracy: 0.9336
Epoch 18/20
accuracy: 0.9756 - val loss: 0.2510 - val accuracy: 0.9362
Epoch 19/20
accuracy: 0.9791 - val loss: 0.2940 - val accuracy: 0.9098
Epoch 20/20
accuracy: 0.9711 - val loss: 0.2712 - val accuracy: 0.9311
```

Transfer Learning using VGG-19 convolutional neural network

Image Augmentation

Fine-tuned Pre-trained VGG-19 CNN Model

Build Model Architecture

```
In [20]: base_vgg_cnn = vgg_cnn
base_vgg_out = base_vgg_cnn.output
vgg_pool_out = tf.keras.layers.Flatten() (base_vgg_out)
hidden1 = tf.keras.layers.Dense(512, activation='relu') (vgg_pool_out)
drop1 = tf.keras.layers.Dropout(rate=0.3) (hidden1)

out = tf.keras.layers.Dense(2, activation='softmax') (drop1)
model_vgg_cnn = tf.keras.Model(inputs=base_vgg_cnn.input, outputs=out)
```

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 125, 125, 3)]	0
olock1_conv1 (Conv2D)	(None, 125, 125, 64)	1792
olock1_conv2 (Conv2D)	(None, 125, 125, 64)	36928
olock1_pool (MaxPooling2D)	(None, 62, 62, 64)	0
olock2_conv1 (Conv2D)	(None, 62, 62, 128)	73856
olock2_conv2 (Conv2D)	(None, 62, 62, 128)	147584
plock2_pool (MaxPooling2D)	(None, 31, 31, 128)	0
plock3_conv1 (Conv2D)	(None, 31, 31, 256)	295168
plock3_conv2 (Conv2D)	(None, 31, 31, 256)	590080
plock3_conv3 (Conv2D)	(None, 31, 31, 256)	590080
plock3_conv4 (Conv2D)	(None, 31, 31, 256)	590080
plock3_pool (MaxPooling2D)	(None, 15, 15, 256)	0
olock4_conv1 (Conv2D)	(None, 15, 15, 512)	1180160
plock4_conv2 (Conv2D)	(None, 15, 15, 512)	2359808
plock4_conv3 (Conv2D)	(None, 15, 15, 512)	2359808
olock4_conv4 (Conv2D)	(None, 15, 15, 512)	2359808
plock4_pool (MaxPooling2D)	(None, 7, 7, 512)	0
olock5_conv1 (Conv2D)	(None, 7, 7, 512)	2359808
plock5_conv2 (Conv2D)	(None, 7, 7, 512)	2359808
olock5_conv3 (Conv2D)	(None, 7, 7, 512)	2359808
plock5_conv4 (Conv2D)	(None, 7, 7, 512)	2359808
plock5_pool (MaxPooling2D)	(None, 3, 3, 512)	0
Flatten_1 (Flatten)	(None, 4608)	0
dense_2 (Dense)	(None, 512)	2359808
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 2)	1026

Non-trainable params: 2,325,568

Train the Model

```
WARNING:tensorflow:sample weight modes were coerced from
  to
 ['...']
WARNING:tensorflow:sample weight modes were coerced from
  to
 ['...']
Train for 271 steps, validate for 30 steps
Epoch 1/20
racy: 0.9205 - val_loss: 0.1241 - val_accuracy: 0.9573
Epoch 2/20
271/271 [============= ] - 150s 555ms/step - loss: 0.1375 - accu
racy: 0.9535 - val loss: 0.1139 - val accuracy: 0.9589
Epoch 3/20
racy: 0.9580 - val loss: 0.1059 - val accuracy: 0.9620
Epoch 4/20
racy: 0.9605 - val_loss: 0.0999 - val_accuracy: 0.9630
Epoch 5/20
racy: 0.9617 - val loss: 0.1068 - val accuracy: 0.9609
Epoch 6/20
racy: 0.9638 - val loss: 0.0939 - val_accuracy: 0.9651
Epoch 7/20
racy: 0.9640 - val loss: 0.0932 - val accuracy: 0.9661
Epoch 8/20
271/271 [============ ] - 1277s 5s/step - loss: 0.1026 - accura
cy: 0.9658 - val loss: 0.0899 - val accuracy: 0.9656
Epoch 9/20
271/271 [============= ] - 223s 824ms/step - loss: 0.0993 - accu
racy: 0.9653 - val loss: 0.0845 - val accuracy: 0.9661
Epoch 10/20
271/271 [============] - 212s 784ms/step - loss: 0.0960 - accu
racy: 0.9664 - val_loss: 0.1028 - val_accuracy: 0.9625
Epoch 11/20
271/271 [============ ] - 212s 781ms/step - loss: 0.0960 - accu
racy: 0.9678 - val loss: 0.0916 - val accuracy: 0.9672
Epoch 12/20
271/271 [============= ] - 215s 792ms/step - loss: 0.0934 - accu
racy: 0.9677 - val loss: 0.0904 - val accuracy: 0.9656
Epoch 13/20
racy: 0.9695 - val loss: 0.0958 - val accuracy: 0.9661
Epoch 14/20
racy: 0.9688 - val_loss: 0.1058 - val_accuracy: 0.9609
Epoch 15/20
racy: 0.9702 - val_loss: 0.1022 - val_accuracy: 0.9646
Epoch 16/20
271/271 [============= ] - 215s 793ms/step - loss: 0.0903 - accu
racy: 0.9692 - val loss: 0.1052 - val_accuracy: 0.9599
Epoch 17/20
271/271 [============== ] - 216s 798ms/step - loss: 0.0841 - accu
racy: 0.9697 - val loss: 0.0779 - val accuracy: 0.9703
Epoch 18/20
racy: 0.9706 - val_loss: 0.0968 - val_accuracy: 0.9641
Epoch 19/20
```

Model Evaluation for Basic CNN and VGG-19 CNN

Retrieve Training and Validation Dataset Accuracy and Loss

```
In [23]: #accuracy
#for Basic CNN
train_accuracy_cnn = history_bsic_cnn.history['accuracy']
validation_accuracy_cnn = history_bsic_cnn.history['val_accuracy']

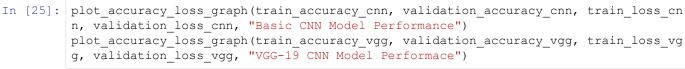
#for VGG-19 cnn
train_accuracy_vgg = history_vgg_cnn.history['accuracy']
validation_accuracy_vgg = history_vgg_cnn.history['val_accuracy']

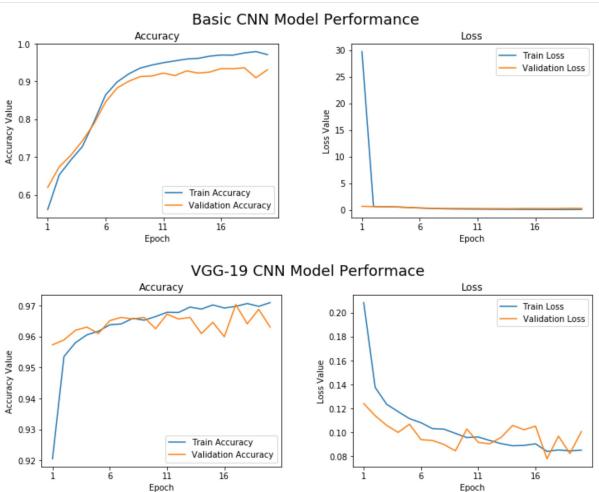
#loss
#for Basic CNN
train_loss_cnn = history_bsic_cnn.history['loss']
validation_loss_cnn = history_bsic_cnn.history['val_loss']

#for VGG-19 CNN
train_loss_vgg = history_vgg_cnn.history['loss']
validation_loss_vgg = history_vgg_cnn.history['val_loss']
```

Plot Accuracy and Loss Graph for Basic CNN and VGG-19 CNN Models

```
In [24]: def plot accuracy loss graph (train accuracy, validation accuracy, train loss, valid
         ation loss, title):
             f, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
             t = f.suptitle(title, fontsize=18)
             f.subplots adjust(top=0.85, wspace=0.3)
             epoch range = len(train accuracy)+1
             epoch list = list(range(1,epoch range))
             ax1.plot(epoch list, train accuracy, label='Train Accuracy')
             ax1.plot(epoch_list, validation_accuracy, label='Validation Accuracy')
             ax1.set xticks(np.arange(1, epoch range, 5))
             ax1.set ylabel('Accuracy Value')
             ax1.set xlabel('Epoch')
             ax1.set_title('Accuracy')
             11 = ax1.legend(loc="best")
             ax2.plot(epoch list, train loss, label='Train Loss')
             ax2.plot(epoch list, validation loss, label='Validation Loss')
             ax2.set_xticks(np.arange(1, epoch_range, 5))
             ax2.set_ylabel('Loss Value')
             ax2.set xlabel('Epoch')
             ax2.set_title('Loss')
             12 = ax2.legend(loc="best")
```





Test Data Accuracy for Basic CNN and VGG-19 CNN Model

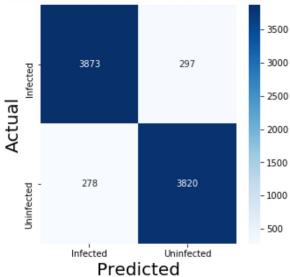
Confusion Matrix for Basic CNN and VGG-19 CNN Model

```
In [31]: from sklearn.metrics import confusion matrix, classification report, accuracy scor
         import seaborn as sns
         def evaluation_matrix(X_test, y_test, model, title):
             pred = model.predict(X test)
             pred = np.argmax(pred,axis = 1)
             y_true = np.argmax(y_test,axis = 1)
             fig = plt.figure(figsize=(5,5))
             ax = fig.gca()
             sns.heatmap(confusion_matrix(y_true , pred),ax=ax,cmap='Blues',annot=True,fmt='
         g',
                        xticklabels = ['Infected', 'Uninfected'],
                        yticklabels=['Infected','Uninfected'])
             ax.set_ylabel('Actual', fontsize=20)
             ax.set xlabel('Predicted', fontsize=20)
             plt.title(title,fontsize=24)
             plt.show()
             print('\n\n\) \n{} '.format(classification\_report(y\_true , pred) ,
                                     accuracy score(y true , pred)))
             plt.figure(1, figsize = (15, 9))
             n = 0
             for i in range (25):
                 n += 1
                 r = np.random.randint( 0 , X_test.shape[0] , 1)
                 plt.subplot(5 , 5 , n)
                 plt.subplots_adjust(hspace = 0.5 , wspace = 0.5)
                 plt.imshow(X test[r[0]])
                 plt.title('true {} : pred {}'.format(y_true[r[0]] , pred[r[0]]) )
                 plt.xticks([]) , plt.yticks([])
             plt.show()
```

Basic CNN Model

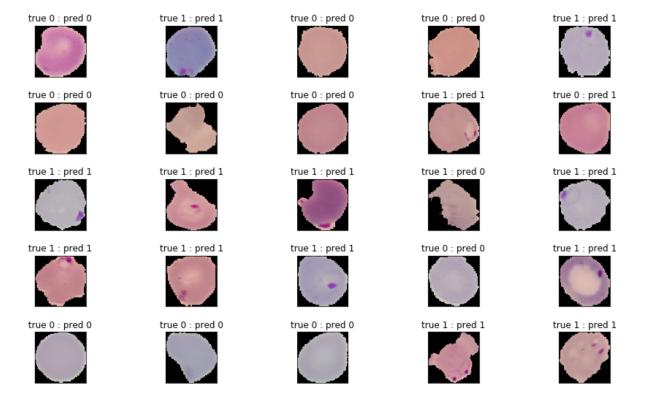
In [32]: evaluation_matrix(X_test, y_test, model_cnn, 'Confustion Matrix for Basic CNN Model
')

Confustion Matrix for Basic CNN Model



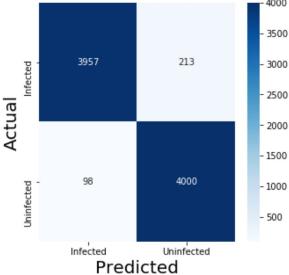
	precision	recall	f1-score	support
0 1	0.93	0.93 0.93	0.93 0.93	4170 4098
accuracy			0.93	8268
macro avg	0.93	0.93	0.93	8268
weighted avg	0.93	0.93	0.93	8268

0.9304547653604257



VGG-19 CNN Model

Confustion Matrix for VGG-19 CNN Model



support	f1-score	recall	precision	
4170	0.96	0.95	0.98	0
4098	0.96	0.98	0.95	1
8268	0.96			accuracy
8268	0.96	0.96	0.96	macro avg
8268	0.96	0.96	0.96	weighted avg

0.962385099177552

