```
In [11]: import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler
         import tensorflow as tf
         from tensorflow.keras.models import Sequential
         from tensorflow.keras.layers import Dense, Dropout, BatchNormalization
         from tensorflow.keras.callbacks import EarlyStopping, ModelCheckpoint
         from tensorflow.keras.metrics import AUC
         # Загрузка данных
         data = pd.read_csv('train_3.2.csv')
         # Основной анализ данных
         print(data.head())
         print(data.describe())
         print(data.isnull().sum())
         print(data['defects'].value_counts())
         # Корреляционная матрица
         plt.figure(figsize=(12, 10))
         sns.heatmap(data.corr(), annot=True, fmt=".2f")
         plt.show()
         # Предобработка данных
         X = data.drop(['id', 'defects'], axis=1)
         y = data['defects'].astype(np.float32)
         # Разделение данных
         X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
         # Нормализация данных
         scaler = StandardScaler()
         X_train_scaled = scaler.fit_transform(X_train)
         X_test_scaled = scaler.transform(X_test)
         # Функция построения базовой модели
         def build_model():
             model = Sequential([
                 Dense(64, activation='relu', input_shape=(X_train_scaled.shape[1]
                 Dense(32, activation='relu'),
                 Dense(1, activation='sigmoid')
             model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[
             return model
         # Обучение модели
         model = build_model()
         history = model.fit(X_train_scaled, y_train, validation_data=(X_test_scal
         # Построение модели с регуляризацией
         def build_regularized_model():
             model = Sequential([
```

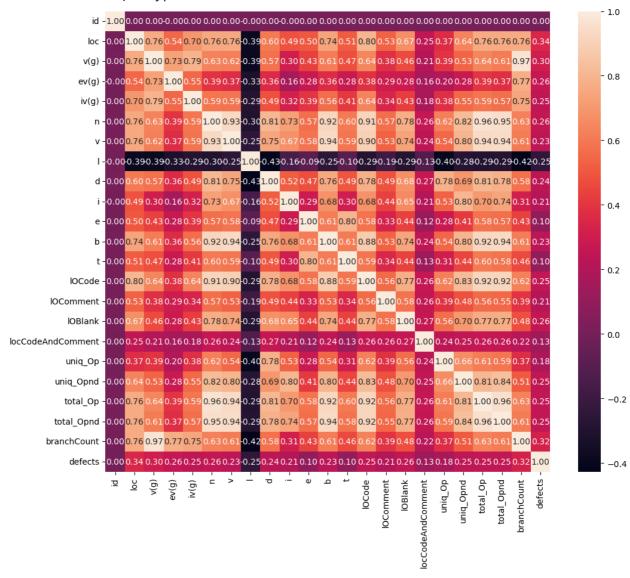
```
Dense(64, activation='relu', input_shape=(X_train_scaled.shape[1]
         BatchNormalization(),
         Dropout(0.5),
         Dense(32, activation='relu'),
         BatchNormalization(),
         Dropout(0.5),
         Dense(1, activation='sigmoid')
     ])
     model.compile(optimizer='adam', loss='binary_crossentropy', metrics=[
     return model
 # Коллбэки
 early_stopping = EarlyStopping(monitor='val_auc', patience=10, mode='max'
 model_checkpoint = ModelCheckpoint('best_model', monitor='val_auc', save_
 # Обучение модели с регуляризацией
 model reg = build regularized model()
 history_reg = model_reg.fit(X_train_scaled, y_train, validation_data=(X_t
 # Загрузка и оценка лучшей модели
 best_model = tf.keras.models.load_model('best_model')
 _, test_auc = best_model.evaluate(X_test_scaled, y test)
 print(f'Test AUC: {test auc}')
 # Визуализация результатов
 plt.plot(history.history['auc'], label='Training AUC (Base Model)')
 plt.plot(history.history['val_auc'], label='Validation AUC (Base Model)')
 plt.plot(history_reg.history['auc'], label='Training AUC (Regularized Mod
 plt.plot(history_reg.history['val_auc'], label='Validation AUC (Regulariz
 plt.title('Model AUC Performance')
 plt.xlabel('Epochs')
 plt.ylabel('AUC')
 plt.legend()
 plt.show()
   id
        loc
             v(g)
                   ev(g)
                           iv(g)
                                                   l
                                                           d
                                                                  i
                                     n
                                             V
                                                                          \
                                                                     . . .
0
    0
       22.0
              3.0
                     1.0
                             2.0
                                  60.0
                                        278.63
                                                0.06
                                                       19.56
                                                              14.25
1
       14.0
              2.0
                     1.0
                             2.0
                                  32.0
                                       151.27
                                                0.14
                                                       7.00
                                                              21.11
2
       11.0
    2
              2.0
                     1.0
                             2.0
                                  45.0
                                       197.65
                                                0.11
                                                       8.05
                                                              22.76
3
        8.0
                     1.0
                             1.0
                                  23.0
                                         94.01
                                                       5.25
                                                              17.86
    3
              1.0
                                                0.19
4
    4
       11.0
              2.0
                     1.0
                             2.0
                                  17.0
                                         60.94 0.18
                                                       5.63 12.44
                                                                     . . .
   10Code
           lOComment lOBlank
                                locCodeAndComment
                                                   uniq_0p uniq_0pnd \
0
       17
                   1
                             1
                                                0
                                                       16.0
                                                                   9.0
1
       11
                   0
                             1
                                                0
                                                       11.0
                                                                  11.0
2
        8
                   0
                             1
                                                0
                                                       12.0
                                                                  11.0
3
        4
                             2
                   0
                                                0
                                                       8.0
                                                                   6.0
4
        7
                   0
                             2
                                                0
                                                       7.0
                                                                   6.0
   total_Op
             total_Opnd branchCount
                                       defects
0
       38.0
                   22.0
                                  5.0
                                         False
1
       18.0
                   14.0
                                  3.0
                                         False
2
       28.0
                   17.0
                                  3.0
                                         False
3
       16.0
                    7.0
                                  1.0
                                          True
4
       10.0
                   10.0
                                  3.0
                                         False
```

```
id
                                   loc
                                                  v(g)
                                                                  ev(g)
       101763.000000
                        101763.000000
                                        101763.000000
                                                         101763.000000
count
mean
        50881.000000
                            37.347160
                                              5,492684
                                                              2.845022
std
        29376.592059
                            54.600401
                                              7.900855
                                                              4.631262
min
             0.000000
                             1.000000
                                              1.000000
                                                              1.000000
25%
        25440.500000
                            13.000000
                                              2.000000
                                                              1.000000
50%
        50881,000000
                            22,000000
                                              3,000000
                                                              1.000000
        76321.500000
75%
                            42.000000
                                              6.000000
                                                              3.000000
       101762.000000
                          3442.000000
                                            404.000000
                                                            165.000000
max
                                                                      ι
                iv(q)
                                                                         \
                                     n
count
       101763.000000
                        101763.000000
                                        101763.000000
                                                         101763.000000
             3.498826
                            96.655995
                                            538.280956
                                                              0.111634
mean
std
             5.534541
                           171.147191
                                          1270.791601
                                                              0.100096
                             0.000000
min
             1.000000
                                              0.000000
                                                              0.000000
25%
             1.000000
                            25.000000
                                             97.670000
                                                              0.050000
50%
             2.000000
                            51.000000
                                           232.790000
                                                              0.090000
75%
             4.000000
                           111.000000
                                            560.250000
                                                              0.150000
                          8441.000000
           402,000000
                                         80843,080000
                                                              1.000000
max
                    d
                                                           t
                                                                      10Code
                                                                               \
                                     i
                                         . . .
       101763.000000
                        101763.000000
                                              101763.000000
                                                              101763.000000
count
                                         . . .
mean
            13.681881
                            27.573007
                                                1141.357982
                                                                   22.802453
std
                                                9862,795472
            14.121306
                            22.856742
                                                                   38.541010
                                         . . .
min
             0.000000
                             0.000000
                                                   0.000000
                                                                    0.000000
25%
             5.600000
                            15.560000
                                                  31.380000
                                                                    7.000000
50%
             9.820000
                            23.360000
                                                 125.400000
                                                                   14.000000
75%
            18.000000
                            34.340000
                                                 565.920000
                                                                   26.000000
           418.200000
                           569.780000
                                              935923.390000
                                                                2824.000000
max
                                         . . .
                                        locCodeAndComment
            10Comment
                               10Blank
                                                                    uniq Op
                                                                              \
                                                             101763.000000
       101763.000000
count
                        101763.000000
                                             101763.000000
                                                  0.196604
             1.773945
                             3.979865
                                                                  11.896131
mean
std
             5.902412
                             6.382358
                                                  0.998906
                                                                   6.749549
                                                  0.000000
                                                                   0.000000
min
             0.000000
                             0.000000
25%
             0.000000
                             1.000000
                                                  0.000000
                                                                   8.000000
50%
             0.000000
                             2.000000
                                                  0.000000
                                                                  11.000000
75%
             1.000000
                             5.000000
                                                  0.000000
                                                                  16.000000
           344.000000
                           219,000000
                                                 43.000000
max
                                                                 410.000000
                                                           branchCount
            uniq_Opnd
                             total_Op
                                           total_Opnd
count
       101763.000000
                        101763.000000
                                        101763.000000
                                                         101763.000000
mean
            15.596671
                            57,628116
                                             39,249698
                                                              9.839549
std
                                             71.692309
            18.064261
                           104.537660
                                                             14,412769
min
             0.000000
                             0.000000
                                              0.000000
                                                              1.000000
25%
             7.000000
                            15.000000
                                             10.000000
                                                              3.000000
50%
            12.000000
                            30.000000
                                             20.000000
                                                              5.000000
75%
            20.000000
                            66.000000
                                             45.000000
                                                             11.000000
          1026,000000
                          5420,000000
                                          3021,000000
max
                                                            503.000000
[8 rows x 22 columns]
id
                       0
                       0
loc
                       0
v(g)
ev(q)
                       0
                       0
iv(g)
                       0
```

n

V	0
1	0
d	0
i	0
е	0
b	0
t	0
10Code	0
lOComment	0
lOBlank	0
locCodeAndComment	0
uniq_Op	0
uniq_Opnd	0
total_Op	0
total_Opnd	0
branchCount	0
defects	0
dtypo: int64	

Name: defects, dtype: int64



```
auc: 0.7878 - val_loss: 0.4312 - val_auc: 0.7864
Epoch 3/100
auc: 0.7886 - val_loss: 0.4333 - val_auc: 0.7841
Epoch 4/100
auc: 0.7893 - val_loss: 0.4311 - val_auc: 0.7866
Epoch 5/100
- auc: 0.7901 - val loss: 0.4313 - val auc: 0.7870
Epoch 6/100
- auc: 0.7899 - val_loss: 0.4315 - val_auc: 0.7854
Epoch 7/100
2545/2545 [============== ] - 3s 1ms/step - loss: 0.4306 -
auc: 0.7904 - val_loss: 0.4309 - val_auc: 0.7868
Epoch 8/100
auc: 0.7907 - val_loss: 0.4303 - val_auc: 0.7868
Epoch 9/100
auc: 0.7905 - val_loss: 0.4296 - val_auc: 0.7865
Epoch 10/100
auc: 0.7913 - val_loss: 0.4297 - val_auc: 0.7870
Epoch 11/100
- auc: 0.7911 - val_loss: 0.4312 - val_auc: 0.7869
Epoch 12/100
2545/2545 [=============== ] - 3s 1ms/step - loss: 0.4293 -
auc: 0.7915 - val_loss: 0.4303 - val_auc: 0.7861
Epoch 13/100
2545/2545 [=============== ] - 3s 992us/step - loss: 0.4294
- auc: 0.7915 - val_loss: 0.4301 - val_auc: 0.7863
Epoch 14/100
auc: 0.7918 - val_loss: 0.4306 - val_auc: 0.7865
Epoch 15/100
auc: 0.7921 - val_loss: 0.4313 - val_auc: 0.7858
Epoch 16/100
auc: 0.7924 - val_loss: 0.4304 - val_auc: 0.7865
Epoch 17/100
auc: 0.7925 - val_loss: 0.4303 - val_auc: 0.7867
Epoch 18/100
auc: 0.7925 - val_loss: 0.4309 - val_auc: 0.7863
Epoch 19/100
- auc: 0.7928 - val_loss: 0.4319 - val_auc: 0.7859
Epoch 20/100
auc: 0.7932 - val_loss: 0.4307 - val_auc: 0.7862
Epoch 21/100
```

```
- auc: 0.7931 - val_loss: 0.4317 - val_auc: 0.7863
Epoch 22/100
auc: 0.7931 - val_loss: 0.4319 - val_auc: 0.7852
Epoch 23/100
- auc: 0.7932 - val_loss: 0.4312 - val_auc: 0.7860
Epoch 24/100
auc: 0.7933 - val loss: 0.4326 - val auc: 0.7862
Epoch 25/100
auc: 0.7938 - val_loss: 0.4321 - val_auc: 0.7849
Epoch 26/100
- auc: 0.7941 - val_loss: 0.4315 - val_auc: 0.7860
Epoch 27/100
auc: 0.7945 - val_loss: 0.4322 - val_auc: 0.7867
Epoch 28/100
auc: 0.7943 - val_loss: 0.4326 - val_auc: 0.7861
Epoch 29/100
- auc: 0.7949 - val_loss: 0.4321 - val_auc: 0.7865
Epoch 30/100
- auc: 0.7948 - val_loss: 0.4333 - val_auc: 0.7851
Epoch 31/100
auc: 0.7953 - val_loss: 0.4316 - val_auc: 0.7854
Epoch 32/100
auc: 0.7957 - val_loss: 0.4343 - val_auc: 0.7855
Epoch 33/100
- auc: 0.7953 - val_loss: 0.4319 - val_auc: 0.7856
Epoch 34/100
- auc: 0.7959 - val_loss: 0.4339 - val_auc: 0.7853
Epoch 35/100
auc: 0.7960 - val_loss: 0.4352 - val_auc: 0.7843
Epoch 36/100
auc: 0.7960 - val_loss: 0.4330 - val_auc: 0.7843
Epoch 37/100
- auc: 0.7965 - val_loss: 0.4348 - val_auc: 0.7838
Epoch 38/100
auc: 0.7966 - val_loss: 0.4337 - val_auc: 0.7852
Epoch 39/100
- auc: 0.7968 - val_loss: 0.4362 - val_auc: 0.7829
Epoch 40/100
```

```
auc: 0.7967 - val_loss: 0.4345 - val_auc: 0.7841
Epoch 41/100
auc: 0.7970 - val_loss: 0.4353 - val_auc: 0.7833
Epoch 42/100
auc: 0.7971 - val_loss: 0.4352 - val_auc: 0.7844
Epoch 43/100
- auc: 0.7972 - val_loss: 0.4345 - val_auc: 0.7845
Epoch 44/100
- auc: 0.7974 - val_loss: 0.4341 - val_auc: 0.7853
Epoch 45/100
2545/2545 [============== ] - 3s 1ms/step - loss: 0.4232 -
auc: 0.7976 - val_loss: 0.4364 - val_auc: 0.7838
Epoch 46/100
- auc: 0.7979 - val_loss: 0.4363 - val_auc: 0.7835
Epoch 47/100
auc: 0.7980 - val_loss: 0.4371 - val_auc: 0.7826
Epoch 48/100
auc: 0.7985 - val_loss: 0.4358 - val_auc: 0.7840
Epoch 49/100
auc: 0.7986 - val_loss: 0.4368 - val_auc: 0.7818
Epoch 50/100
auc: 0.7985 - val_loss: 0.4360 - val_auc: 0.7826
Epoch 51/100
auc: 0.7991 - val_loss: 0.4374 - val_auc: 0.7836
Epoch 52/100
auc: 0.7988 - val_loss: 0.4382 - val_auc: 0.7833
Epoch 53/100
- auc: 0.7993 - val_loss: 0.4378 - val_auc: 0.7825
Epoch 54/100
- auc: 0.7992 - val_loss: 0.4388 - val_auc: 0.7816
Epoch 55/100
auc: 0.7992 - val_loss: 0.4376 - val_auc: 0.7831
Epoch 56/100
auc: 0.7995 - val_loss: 0.4392 - val_auc: 0.7827
Epoch 57/100
- auc: 0.7998 - val_loss: 0.4385 - val_auc: 0.7828
Epoch 58/100
- auc: 0.8001 - val_loss: 0.4382 - val_auc: 0.7823
Epoch 59/100
```

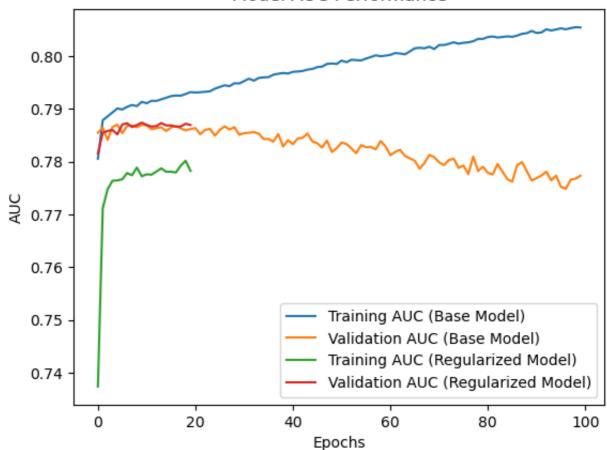
```
- auc: 0.8000 - val_loss: 0.4397 - val_auc: 0.7839
Epoch 60/100
auc: 0.8001 - val_loss: 0.4401 - val_auc: 0.7829
Epoch 61/100
- auc: 0.8002 - val_loss: 0.4399 - val_auc: 0.7812
Epoch 62/100
auc: 0.8006 - val loss: 0.4422 - val auc: 0.7818
Epoch 63/100
auc: 0.8005 - val_loss: 0.4411 - val_auc: 0.7822
Epoch 64/100
2545/2545 [============== ] - 3s 1ms/step - loss: 0.4201 -
auc: 0.8003 - val_loss: 0.4426 - val_auc: 0.7811
Epoch 65/100
auc: 0.8009 - val_loss: 0.4410 - val_auc: 0.7806
Epoch 66/100
auc: 0.8015 - val_loss: 0.4426 - val_auc: 0.7802
Epoch 67/100
auc: 0.8016 - val_loss: 0.4427 - val_auc: 0.7787
Epoch 68/100
- auc: 0.8015 - val_loss: 0.4425 - val_auc: 0.7797
Epoch 69/100
auc: 0.8018 - val_loss: 0.4407 - val_auc: 0.7813
Epoch 70/100
- auc: 0.8013 - val_loss: 0.4398 - val_auc: 0.7809
Epoch 71/100
auc: 0.8021 - val_loss: 0.4404 - val_auc: 0.7799
Epoch 72/100
auc: 0.8021 - val_loss: 0.4445 - val_auc: 0.7793
Epoch 73/100
auc: 0.8023 - val_loss: 0.4417 - val_auc: 0.7803
Epoch 74/100
- auc: 0.8026 - val_loss: 0.4429 - val_auc: 0.7806
Epoch 75/100
- auc: 0.8023 - val_loss: 0.4437 - val_auc: 0.7788
Epoch 76/100
- auc: 0.8025 - val_loss: 0.4449 - val_auc: 0.7792
Epoch 77/100
- auc: 0.8026 - val_loss: 0.4454 - val_auc: 0.7776
Epoch 78/100
```

```
auc: 0.8028 - val_loss: 0.4440 - val_auc: 0.7810
Epoch 79/100
auc: 0.8033 - val_loss: 0.4457 - val_auc: 0.7782
Epoch 80/100
- auc: 0.8032 - val_loss: 0.4481 - val_auc: 0.7790
Epoch 81/100
- auc: 0.8036 - val loss: 0.4479 - val auc: 0.7779
Epoch 82/100
auc: 0.8037 - val_loss: 0.4447 - val_auc: 0.7776
Epoch 83/100
2545/2545 [============== ] - 3s 1ms/step - loss: 0.4169 -
auc: 0.8035 - val_loss: 0.4478 - val_auc: 0.7795
Epoch 84/100
auc: 0.8036 - val_loss: 0.4458 - val_auc: 0.7781
Epoch 85/100
auc: 0.8037 - val_loss: 0.4487 - val_auc: 0.7767
Epoch 86/100
- auc: 0.8036 - val_loss: 0.4492 - val_auc: 0.7762
Epoch 87/100
auc: 0.8039 - val_loss: 0.4467 - val_auc: 0.7793
Epoch 88/100
auc: 0.8042 - val_loss: 0.4500 - val_auc: 0.7799
Epoch 89/100
auc: 0.8043 - val_loss: 0.4494 - val_auc: 0.7782
Epoch 90/100
auc: 0.8048 - val_loss: 0.4532 - val_auc: 0.7764
Epoch 91/100
auc: 0.8044 - val_loss: 0.4542 - val_auc: 0.7769
Epoch 92/100
- auc: 0.8045 - val_loss: 0.4498 - val_auc: 0.7774
Epoch 93/100
auc: 0.8051 - val_loss: 0.4478 - val_auc: 0.7781
Epoch 94/100
auc: 0.8049 - val_loss: 0.4522 - val_auc: 0.7764
Epoch 95/100
auc: 0.8050 - val_loss: 0.4536 - val_auc: 0.7773
Epoch 96/100
auc: 0.8053 - val_loss: 0.4493 - val_auc: 0.7752
Epoch 97/100
```

```
auc: 0.8051 - val_loss: 0.4492 - val_auc: 0.7748
Epoch 98/100
auc: 0.8053 - val_loss: 0.4535 - val_auc: 0.7766
Epoch 99/100
auc: 0.8055 - val_loss: 0.4554 - val_auc: 0.7768
Epoch 100/100
- auc: 0.8054 - val loss: 0.4536 - val auc: 0.7773
Epoch 1/100
0.7372INFO:tensorflow:Assets written to: best_model/assets
INFO:tensorflow:Assets written to: best model/assets
auc: 0.7374 - val loss: 0.4417 - val auc: 0.7815
Epoch 2/100
0.7713INFO:tensorflow:Assets written to: best_model/assets
INFO:tensorflow:Assets written to: best_model/assets
auc: 0.7712 - val_loss: 0.4391 - val_auc: 0.7855
Epoch 3/100
0.7747INFO:tensorflow:Assets written to: best_model/assets
INFO:tensorflow:Assets written to: best_model/assets
auc: 0.7748 - val_loss: 0.4362 - val_auc: 0.7858
Epoch 4/100
0.7765INFO:tensorflow:Assets written to: best model/assets
INFO:tensorflow:Assets written to: best_model/assets
auc: 0.7764 - val_loss: 0.4362 - val_auc: 0.7860
Epoch 5/100
auc: 0.7764 - val_loss: 0.4339 - val_auc: 0.7852
Epoch 6/100
0.7767INFO:tensorflow:Assets written to: best_model/assets
INFO:tensorflow:Assets written to: best_model/assets
auc: 0.7766 - val_loss: 0.4341 - val_auc: 0.7870
Epoch 7/100
0.7780INFO:tensorflow:Assets written to: best_model/assets
INFO:tensorflow:Assets written to: best_model/assets
```

```
auc: 0.7778 - val_loss: 0.4331 - val_auc: 0.7873
Epoch 8/100
auc: 0.7774 - val_loss: 0.4332 - val_auc: 0.7866
Epoch 9/100
auc: 0.7789 - val_loss: 0.4329 - val_auc: 0.7869
Epoch 10/100
0.7770INFO:tensorflow:Assets written to: best model/assets
INFO:tensorflow:Assets written to: best_model/assets
auc: 0.7772 - val_loss: 0.4315 - val_auc: 0.7874
Epoch 11/100
auc: 0.7776 - val loss: 0.4338 - val auc: 0.7869
Epoch 12/100
auc: 0.7775 - val_loss: 0.4333 - val_auc: 0.7866
Epoch 13/100
auc: 0.7781 - val_loss: 0.4320 - val_auc: 0.7867
Epoch 14/100
auc: 0.7787 - val_loss: 0.4317 - val_auc: 0.7873
Epoch 15/100
auc: 0.7781 - val loss: 0.4333 - val auc: 0.7869
Epoch 16/100
auc: 0.7781 - val_loss: 0.4337 - val_auc: 0.7869
Epoch 17/100
auc: 0.7779 - val_loss: 0.4322 - val_auc: 0.7866
Epoch 18/100
auc: 0.7792 - val_loss: 0.4329 - val_auc: 0.7867
Epoch 19/100
auc: 0.7802 - val loss: 0.4350 - val auc: 0.7872
Epoch 20/100
auc: 0.7782 - val_loss: 0.4320 - val_auc: 0.7869
auc: 0.7874
Test AUC: 0.7873856425285339
```

Model AUC Performance



In []: