Knee Osteoarthritis Classification using Federated Learning with FedAvg Approach

IMPLEMENTATION RESULTS

1. DenseNet169 - TensorFlow

Global Model Accuracy: 83.79 %

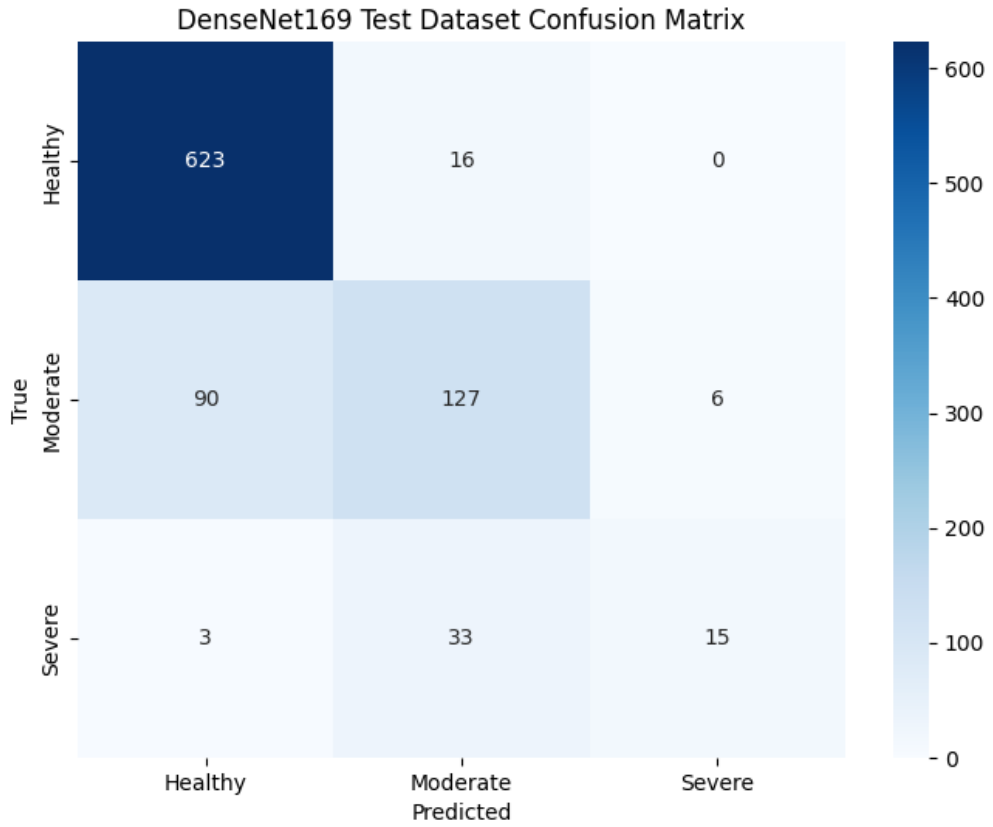
Global Model Loss: 0.4182

```
Final Test Results:
--------------------
Test Loss: 0.4182
Test Accuracy: 0.8379

AutoTest Results:
--------------------
Test Loss: 0.3786
Test Accuracy: 0.8479
29/29 [==============================] - 21s 638ms/step

Test dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.87      0.97      0.92       639
    Moderate       0.72      0.57      0.64       223
      Severe       0.71      0.29      0.42        51

    accuracy                           0.84       913
   macro avg       0.77      0.61      0.66       913
weighted avg       0.83      0.84      0.82       913

27/27 [==============================] - 17s 607ms/step

Autotest dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.88      0.98      0.92       604
    Moderate       0.73      0.57      0.64       200
      Severe       0.78      0.32      0.45        44

    accuracy                           0.85       848
   macro avg       0.79      0.62      0.67       848
weighted avg       0.84      0.85      0.83       848
```

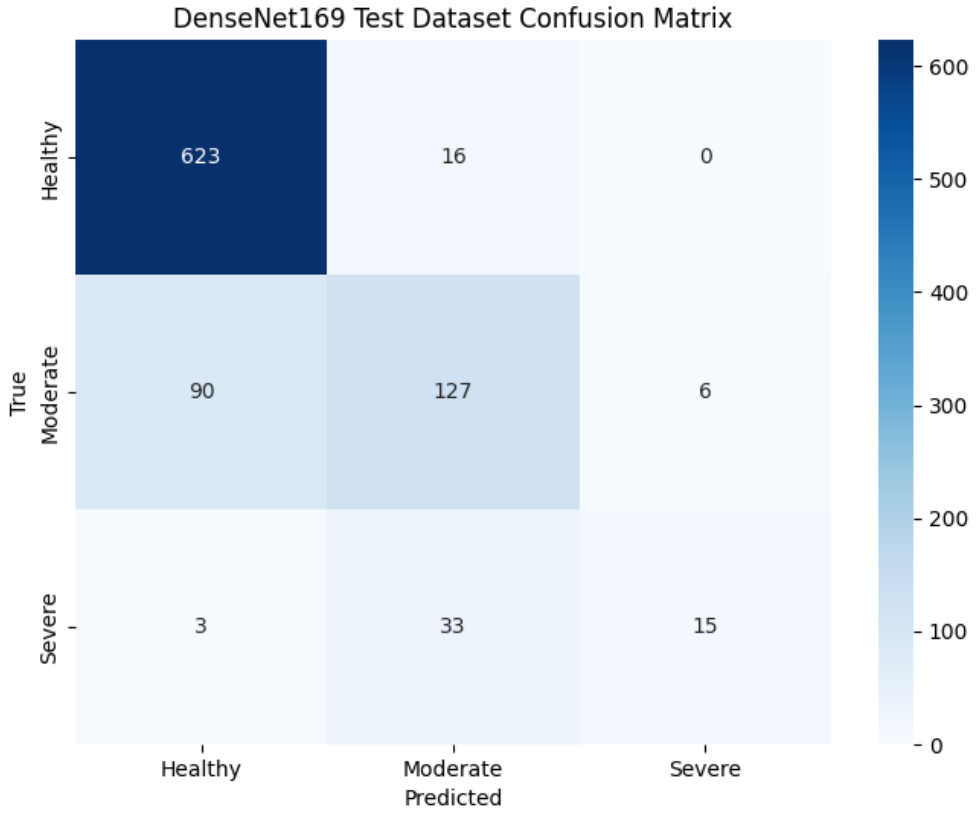
Round Accuracies and Losses:

Round	Global Accuracy
1	73.10 %
2	78.95 %
3	80.26 %
4	82.86 %
5	83.51 %
6	83.51 %
7	84.81 %
8	83.51 %
9	84.81 %
10	83.94 %
11	84.59 %
12	85.68 %

Test Dataset Confusion Matrix:



Auto-test Dataset Confusion Matrix:



2. **DenseNet201 Model – TensorFlow Framework:**

Global Model Accuracy: 83.79 %

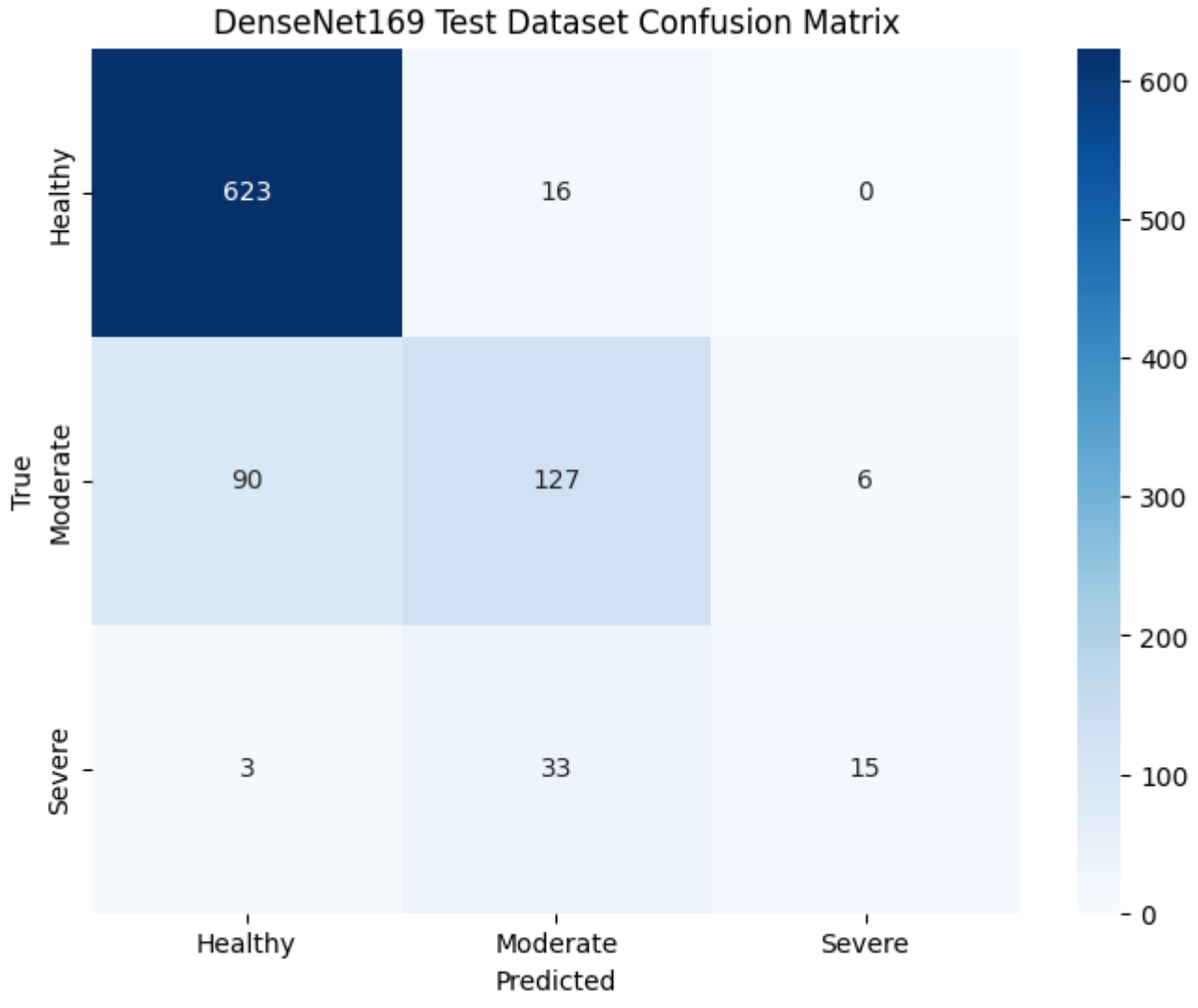
Global Model Loss: 0.4182

```
Final Test Results:
-------------------
Test Loss: 0.3518
Test Accuracy: 0.8510

AutoTest Results:
-------------------
Test Loss: 0.3322
Test Accuracy: 0.8703
29/29 [==============================] - 27s 818ms/step

Test dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.89      0.98      0.93       639
    Moderate       0.79      0.55      0.65       223
      Severe       0.52      0.59      0.55        51

    accuracy                           0.85       913
   macro avg       0.73      0.70      0.71       913
weighted avg       0.85      0.85      0.84       913

27/27 [==============================] - 22s 801ms/step

Autotest dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.90      0.99      0.94       604
    Moderate       0.85      0.56      0.68       200
      Severe       0.57      0.70      0.63        44

    accuracy                           0.87       848
   macro avg       0.78      0.75      0.75       848
weighted avg       0.87      0.87      0.86       848
```

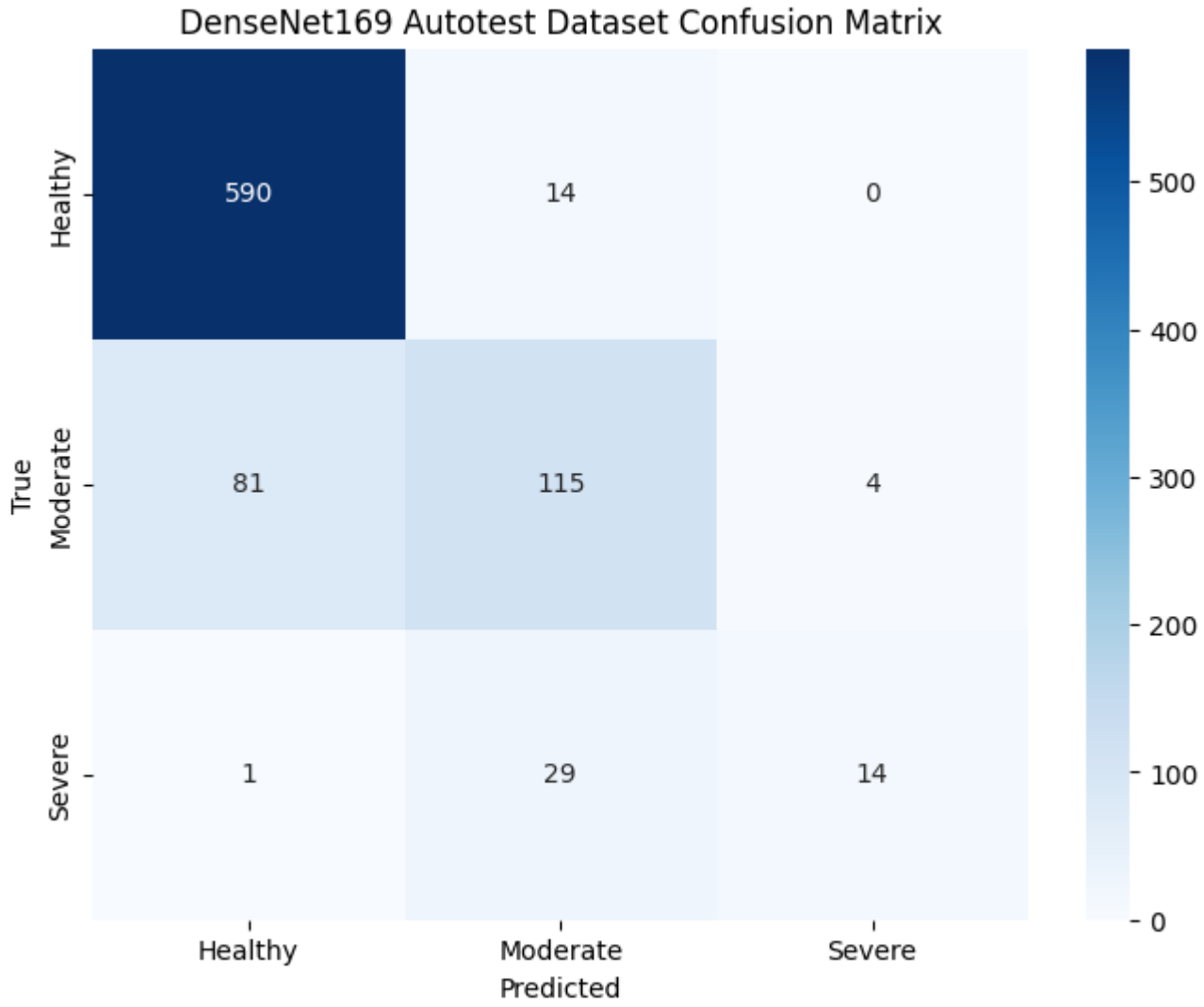
Round Accuracies and Losses:

Round	Global Accuracy
1	73.10 %
2	78.96 %
3	80.26 %
4	82.86 %
5	83.51 %
6	83.51 %
7	84.82 %
8	83.51 %
9	84.82 %
10	83.95 %
11	84.60 %
12	85.68 %

Test Dataset Confusion Matrix:



Auto-test Confusion Matrix:



3. InceptionResNetV2 Model – TensorFlow Framework

Global Model Accuracy: 79.41 %

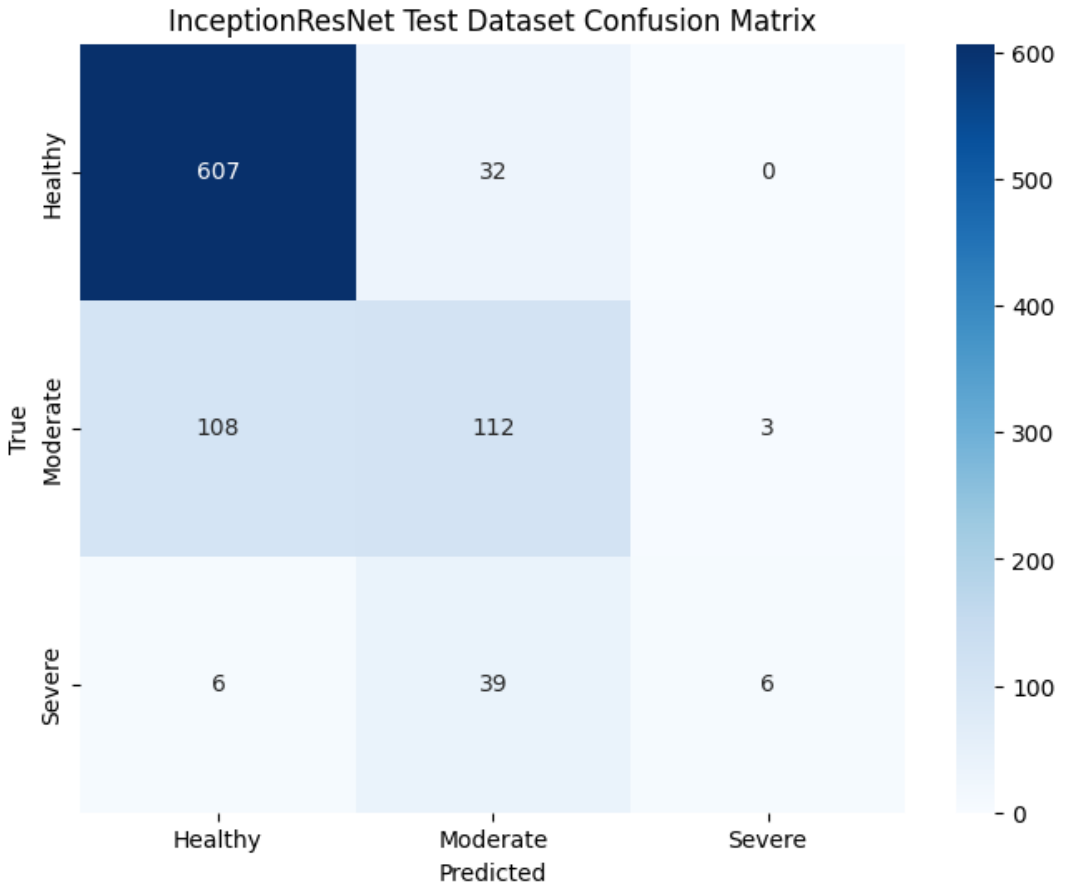
Global Model Loss: 0.4760

```
Final Test Results:
--------------------
Test Loss: 0.4760
Test Accuracy: 0.7941

AutoTest Results:
--------------------
Test Loss: 0.4474
Test Accuracy: 0.8184
29/29 [==============================] - 17s 475ms/step

Test dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.84      0.95      0.89       639
    Moderate       0.61      0.50      0.55       223
      Severe       0.67      0.12      0.20        51

    accuracy                           0.79       913
   macro avg       0.71      0.52      0.55       913
weighted avg       0.78      0.79      0.77       913

27/27 [==============================] - 13s 464ms/step

Autotest dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.86      0.97      0.91       604
    Moderate       0.65      0.54      0.59       200
      Severe       1.00      0.07      0.13        44

    accuracy                           0.82       848
   macro avg       0.84      0.52      0.54       848
weighted avg       0.82      0.82      0.79       848
```

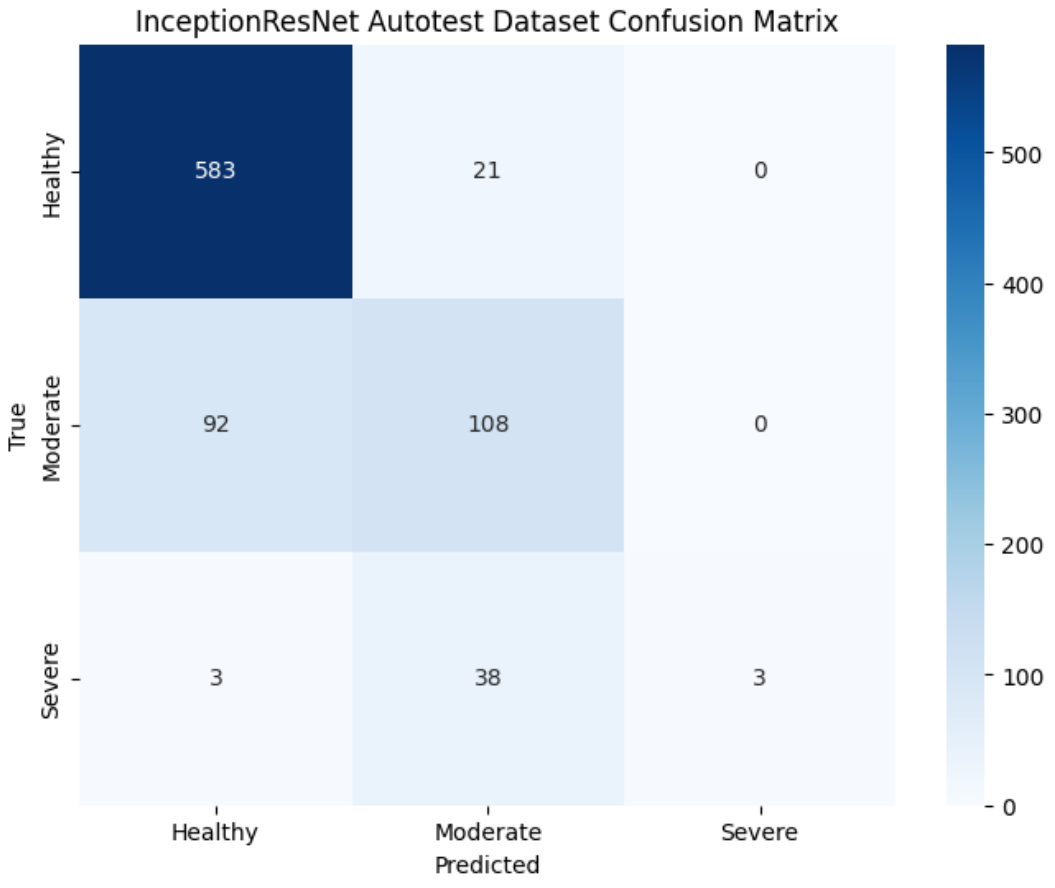
Round Accuracies and Losses:

Round	Global Accuracy
1	73.97 %
2	77.01 %
3	76.79 %
4	76.79 %
5	79.18 %
6	79.18 %
7	79.39 %
8	79.39 %
9	80.91 %
10	82.00 %
11	81.13 %
12	82.65 %

Test Dataset Confusion Matrix:



Auto-Test Confusion Matrix:



4. DenseNet201 Model – PyTorch Framework

Global Model Accuracy: 95.18 %

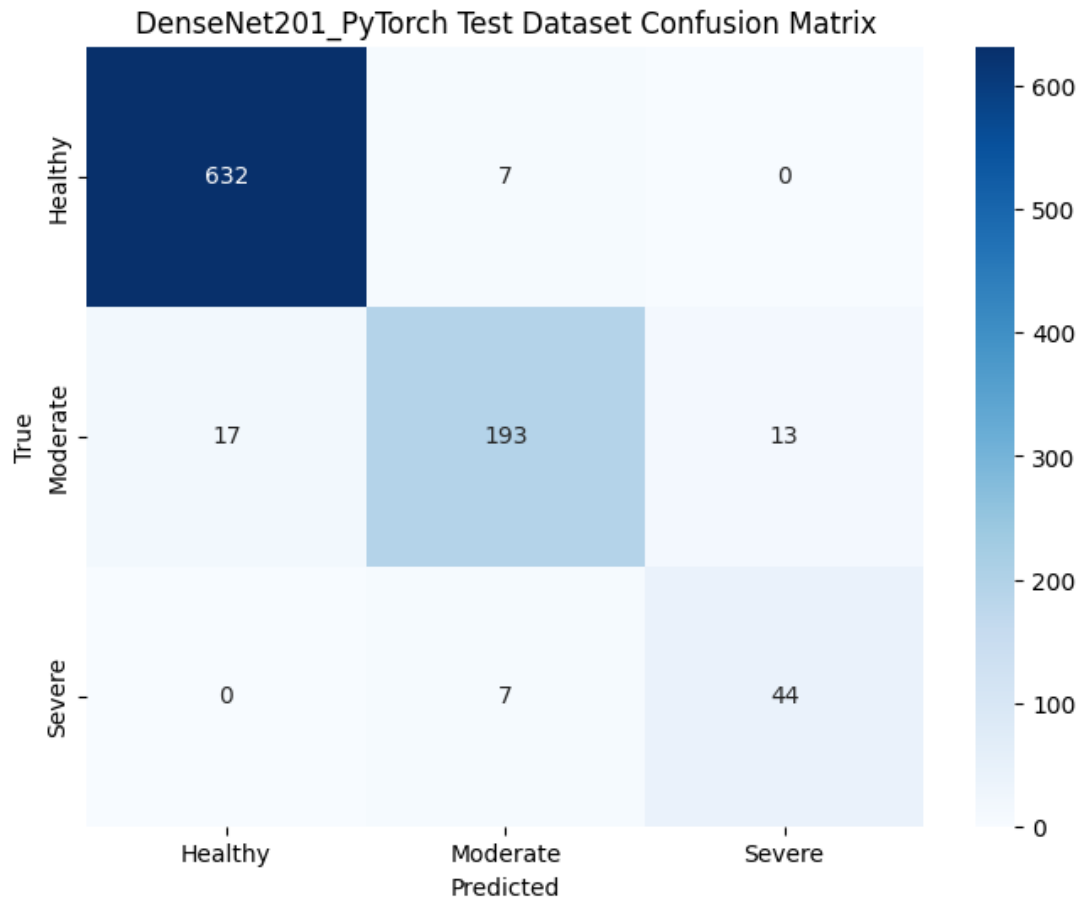
Global Model Loss: 0.1455

```
Final Test Results:
--------------------
Test Accuracy: 0.9518
Test Loss: 0.1455

AutoTest Results:
--------------------
AutoTest Accuracy: 0.9587
AutoTest Loss: 0.1246

Test dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.97      0.99      0.98       639
    Moderate       0.93      0.87      0.90       223
      Severe       0.77      0.86      0.81        51

    accuracy                           0.95       913
   macro avg       0.89      0.91      0.90       913
weighted avg       0.95      0.95      0.95       913


Autotest dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.97      0.99      0.98       604
    Moderate       0.96      0.86      0.91       200
      Severe       0.78      0.91      0.84        44

    accuracy                           0.96       848
   macro avg       0.90      0.92      0.91       848
weighted avg       0.96      0.96      0.96       848
```

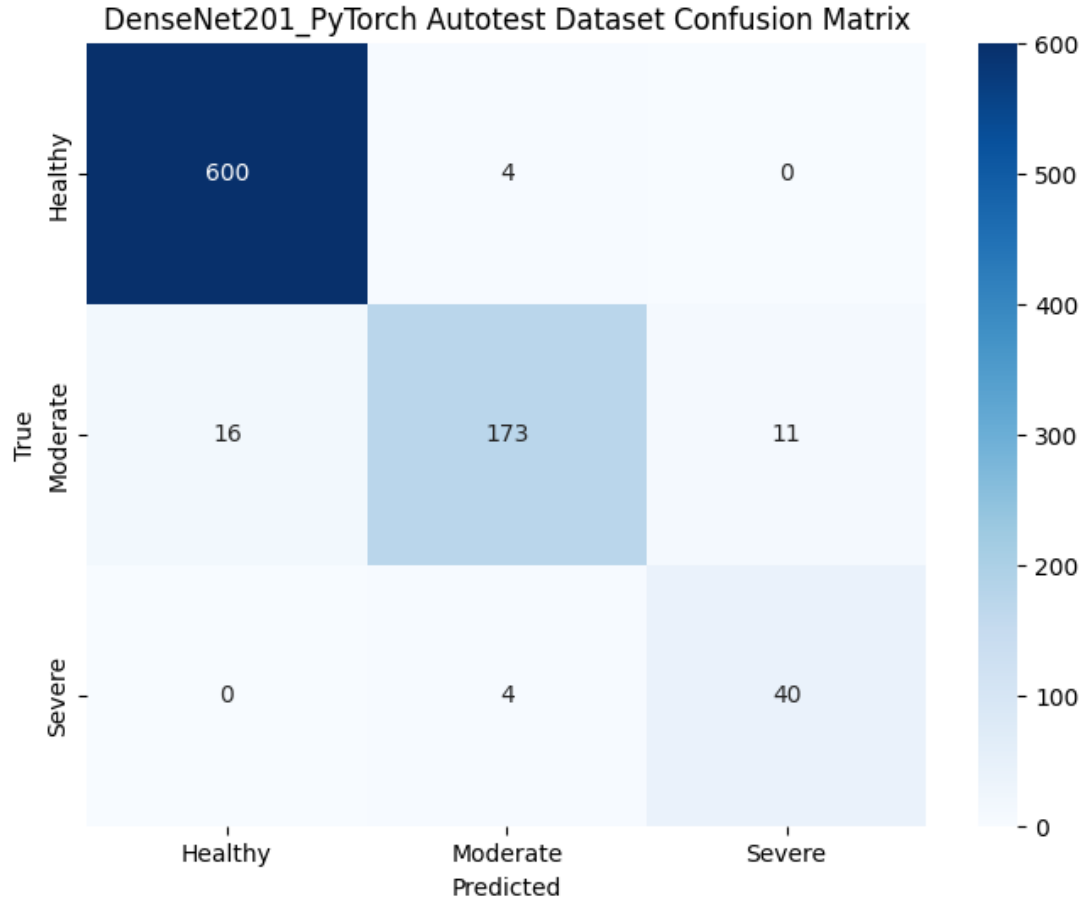
Round Accuracies and Losses:

Round	Global Accuracy
1	71.15 %
2	84.38 %
3	86.33 %
4	86.12 %
5	90.67 %
6	90.46 %
7	91.32 %
8	92.62 %
9	94.14 %
10	93.71 %
11	90.89 %
12	93.28 %

Test Dataset Confusion Matrix:



Auto-Test Confusion Matrix:



5. InceptionResNetV2 Model – PyTorch Framework:

Global Model Accuracy: 94.58 %

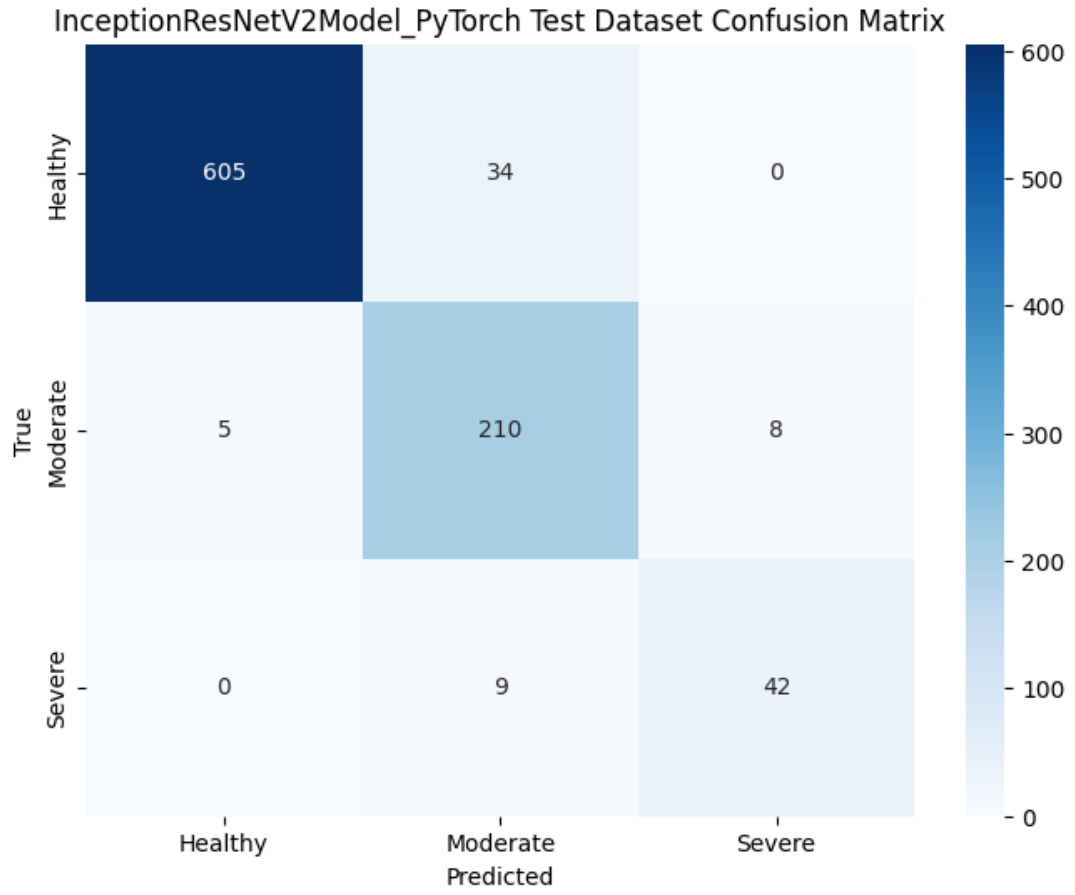
Global Model Loss: 0.1813

```
Final Test Results:
--------------------
Test Accuracy: 0.9387
Test Loss: 0.1813

AutoTest Results:
--------------------
AutoTest Accuracy: 0.9458
AutoTest Loss: 0.1539

Test dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       0.99      0.95      0.97       639
    Moderate       0.83      0.94      0.88       223
      Severe       0.84      0.82      0.83        51

    accuracy                           0.94       913
   macro avg       0.89      0.90      0.89       913
weighted avg       0.94      0.94      0.94       913


Autotest dataset - Classification Report:
              precision    recall  f1-score   support

     Healthy       1.00      0.95      0.97       604
    Moderate       0.84      0.95      0.89       200
      Severe       0.84      0.84      0.84        44

    accuracy                           0.95       848
   macro avg       0.89      0.92      0.90       848
weighted avg       0.95      0.95      0.95       848
```

Round Accuracies and Losses:

Round	Global Accuracy
1	85.68 %
2	88.07 %
3	88.29 %
4	91.32 %
5	92.62 %
6	93.49 %
7	94.14 %
8	93.93 %
9	93.93 %
10	94.14 %
11	95.23 %
12	91.11 %

Test Dataset Confusion Matrix:



Auto-Test Dataset Matrix: