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

     Moderate
                                                     0.42

      accuracy
      0.84
      913

      macro avg
      0.77
      0.61
      0.66
      913

      ighted avg
      0.83
      0.84
      0.82
      913

weighted avg
27/27 [=========== ] - 17s 607ms/step
Autotest dataset - Classification Report:
                   precision recall f1-score support

    0.88
    0.98
    0.92

    0.73
    0.57
    0.64

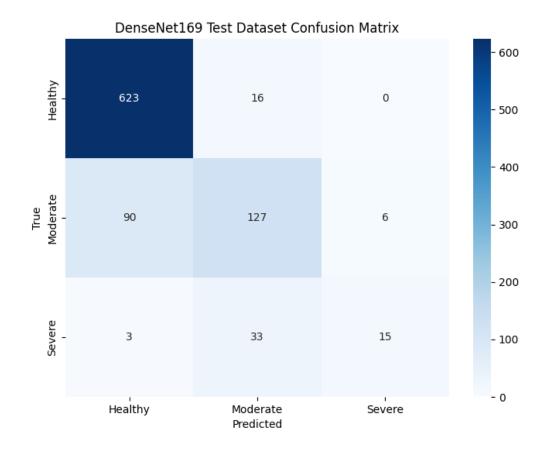
    0.78
    0.32
    0.45

     Healthy
Moderate
Severe
                                                                    604
                                                                     200
                                                                      44
                                                      0.85
                                                                    848
     accuracy
macro avg 0.79 0.62 0.67
weighted avg 0.84 0.85 0.83
                                                                     848
                                                                       848
```

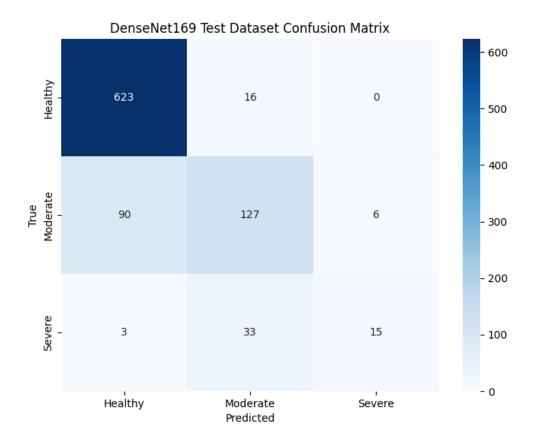
Round Accuracies and Losses:

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

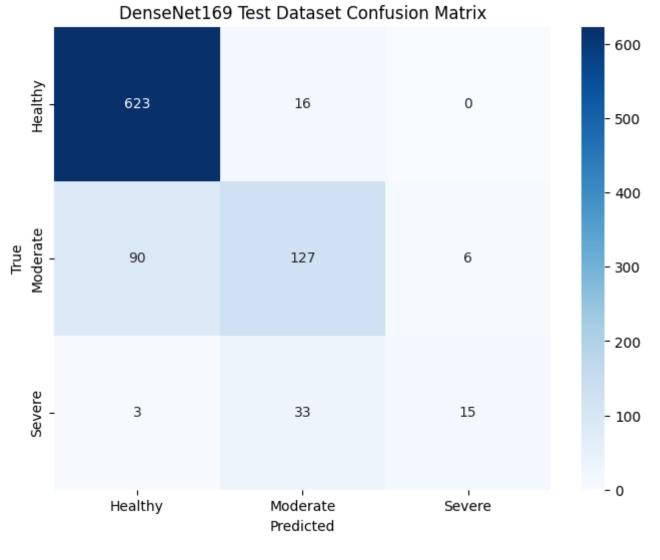
Global Model Li	0.4102			
Final Test Results:				
Test Loss: 0.	3518			
Test Accuracy	: 0.8510			
AutoTest Resu	ılts:			
Test Loss: 0.	3322			
Test Accuracy	: 0.8703			
29/29 [=====			===1 - 27s	818ms/step
23,23 [, 2.3	этэшэ, этгр
Test dataset	- Classifica	ation Reno	rt.	
rest uataset	precision			sunnort
	bi ectatori	recarr	11-30016	зиррог с
Healthy	0.89	0.98	0.93	639
_	0.79			
	0.52			
Severe	0.32	0.55	0.55	71
accuracy			0.85	913
macro avg		0.70	0.71	
weighted avg				
nerbucea arb	0.03	0.03	0.01	313
27/27 [=====			1 - 226	901mc/cton
21/21 [223	oomiis/ arch
Autotest data	set - Classi	ification	Report:	
nacoccoc data	precision		f1-score	support
	pi cc1310ii	rccarr	11 30010	зиррог с
Healthy	0.90	0.99	0.94	604
Moderate	0.85	0.56	0.68	200
Severe	0.57	0.70	0.63	44
Jevel e	0.57	0.70	0.05	
accuracy			0.87	848
macro avg	0.78	0.75	0.75	848
weighted avg	0.73 0.87	0.73	0.86	848
weighted avg	0.67	0.07	0.00	040

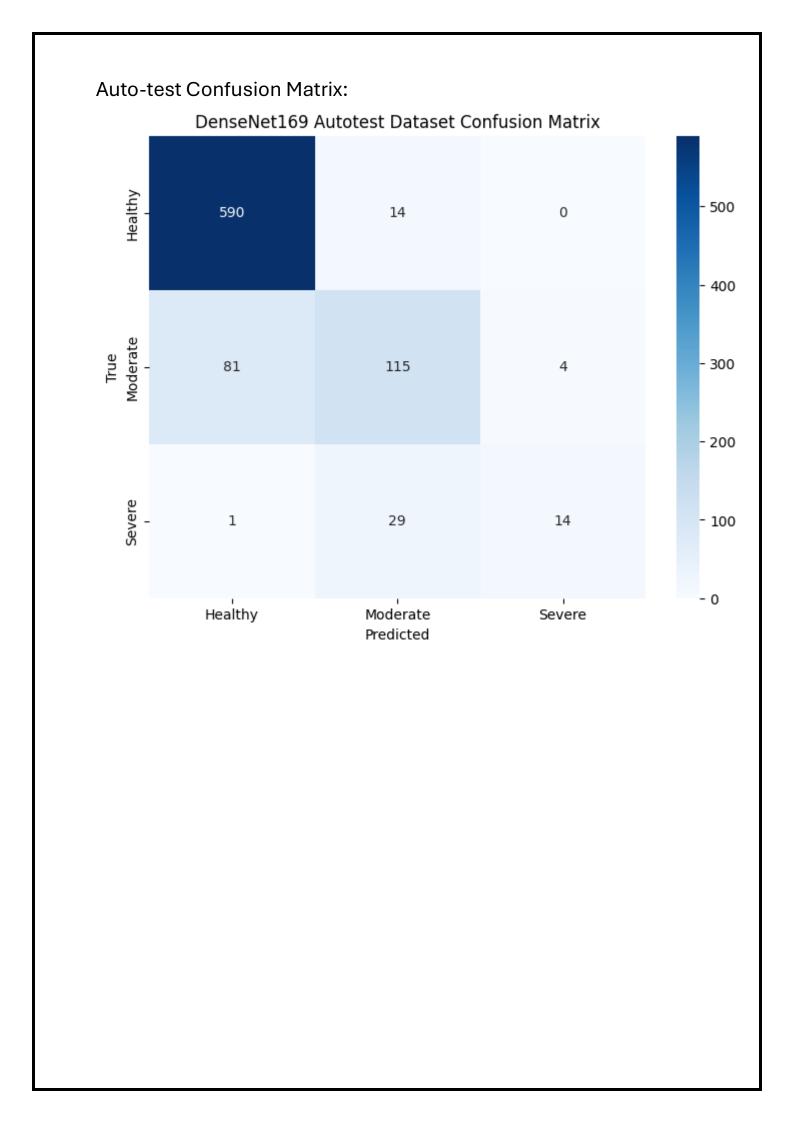
Round Accuracies and Losses:

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:







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

      Moderate
      0.61
      0.50
      0.55

      Severe
      0.67
      0.12
      0.20

                                                          639
                                                          223
                                              0.79
                                                           913
    accuracy
                 0.71 0.52
0.78 0.79
   macro avg
                                              0.55
weighted avg
                                              0.77
                                                            913
27/27 [=====] - 13s 464ms/step
Autotest dataset - Classification Report:
                precision recall f1-score support

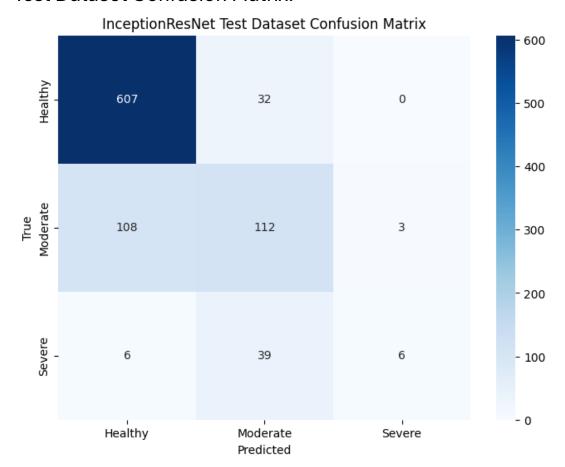
      0.86
      0.97

      0.65
      0.54

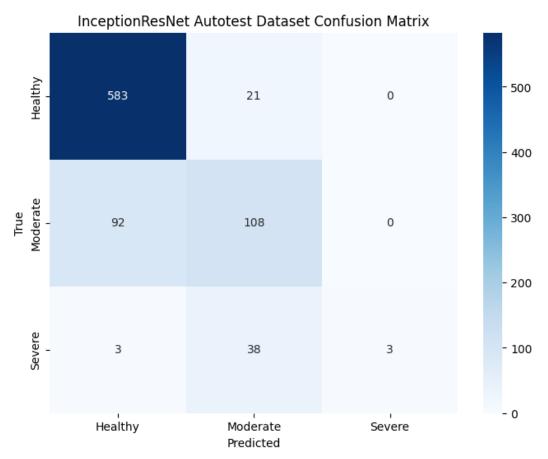
      1.00
      0.07

                                            0.91
0.59
     Healthy
    Moderate
                                                            200
                                            0.13
      Severe
                                                           44
    accuracy
                                              0.82
                                                            848
                      0.84 0.52
   macro avg
                                              0.54
                                                            848
weighted avg
                              0.82
                                              0.79
                      0.82
                                                            848
```

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 %



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

      Moderate
      0.93
      0.87
      0.90

      Severe
      0.77
      0.86
      0.81

    accuracy
                                            0.95 913
macro avg 0.89 0.91 0.90
weighted avg 0.95 0.95 0.95
Autotest dataset - Classification Report:
               precision recall f1-score support
                   0.97 0.99 0.98
     Healthy
                                                            604

    0.96
    0.86
    0.91

    0.78
    0.91
    0.84

    Moderate
                                                            200
       Severe
                                             0.96
                                                            848
    accuracy
                      0.90 0.92
                                            0.91
                                                            848
   macro avg
                    0.96
weighted avg
                                0.96
                                            0.96
                                                            848
```

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: DenseNet201_PyTorch Test Dataset Confusion Matrix 600 Healthy 632 7 0 - 500 - 400 True Moderate ' 17 193 13 - 300 - 200 Severe 7 0 44 - 100

Moderate

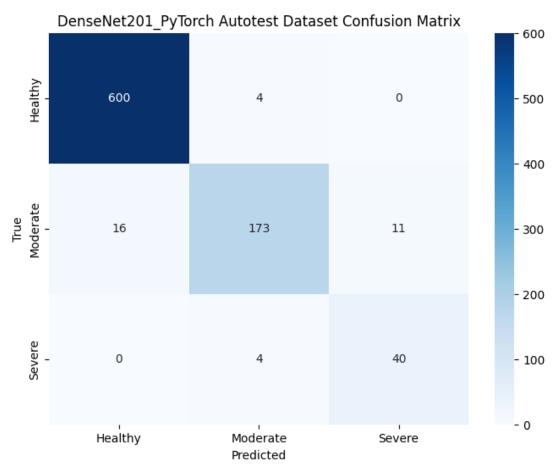
Predicted

- 0

Severe

Auto-Test Confusion Matrix:

Healthy



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

                                               0.94 913
    accuracy
macro avg 0.89 0.90 0.89
weighted avg 0.94 0.94 0.94
                                                             913
                                                                913
Autotest dataset - Classification Report:
                  precision recall f1-score support
                     1.00 0.95 0.97
0.84 0.95 0.89
0.84 0.84 0.84
     Healthy
                                                                604
     Moderate
Severe
                                                               200
                                                                 44
                                                0.95
                                                              848
     accuracy
macro avg 0.89 0.92 0.90 weighted avg 0.95 0.95 0.95
                                                                848
                                                                848
```

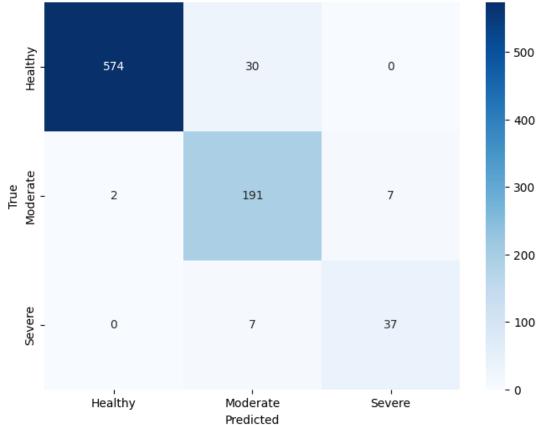
Round	Global Accuracy
1	85.68 %
2	88.07 %
3	88.29 %
4	91.32 %
5	92.62 %
6	93.49 %
7	94.14 %
8	93.93 %
9	93.93 %
10	94.14 %
11	95.23 %
12	91.11 %

InceptionResNetV2Model_PyTorch Test Dataset Confusion Matrix



Auto-Test Dataset Matrix:

InceptionResNetV2Model PyTorch Autotest Dataset Confusion Matrix



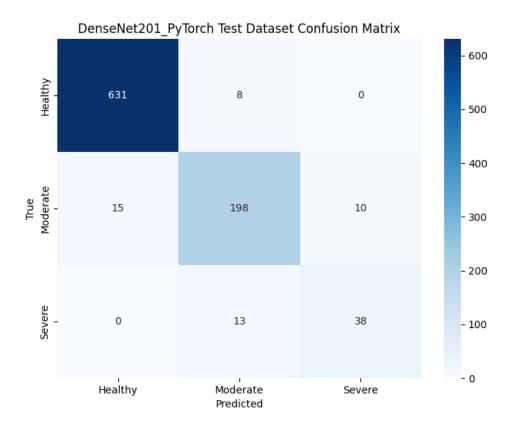
6. DenseNet201 Model - PyTorch Framework - Weight Initialization

Global Model Accuracy: 94.96%

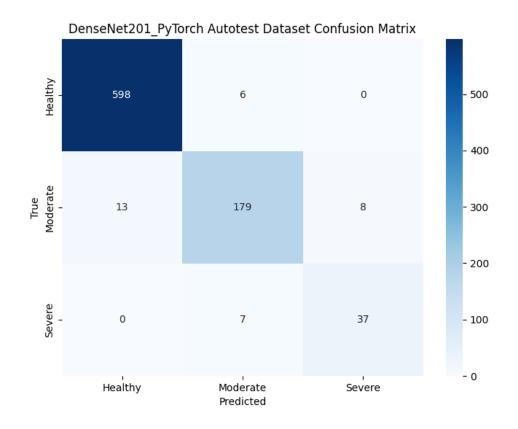
Global Model Loss: 0.1460

Final Test Results:						
Test Loss: 0.1460 AutoTest Results:	Final Test Results:					
Test Loss: 0.1460 AutoTest Results:	Test Accuracy	. 0 9496				
AutoTest Results:	-					
AutoTest Accuracy: 0.9599 AutoTest Loss: 0.1303 Test dataset - Classification Report:	Test Loss. V.	1400				
AutoTest Accuracy: 0.9599 AutoTest Loss: 0.1303 Test dataset - Classification Report:	AutoTost Bosi	ıl+cı				
AutoTest Loss: 0.1303 Test dataset - Classification Report:	Autorest Rest	1113.				
AutoTest Loss: 0.1303 Test dataset - Classification Report:	AutoTest Acci		a			
Test dataset - Classification Report:		-	,			
### Precision recall f1-score support Healthy	Autorest Loss	o. 0.1505				
Healthy 0.98 0.99 0.98 639 Moderate 0.90 0.89 0.90 223 Severe 0.79 0.75 0.77 51 accuracy 0.95 913 macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report: precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	Test dataset	- Classific	ation Repo	rt:		
Healthy 0.98 0.99 0.98 639 Moderate 0.90 0.89 0.90 223 Severe 0.79 0.75 0.77 51 accuracy 0.95 913 macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report:		precision	recall	f1-score	support	
Moderate 0.90 0.89 0.90 223 Severe 0.79 0.75 0.77 51 accuracy 0.95 913 macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report:						
Moderate 0.90 0.89 0.90 223 Severe 0.79 0.75 0.77 51 accuracy 0.95 913 macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report:	Healthy	0.98	0.99	0.98	639	
Severe 0.79 0.75 0.77 51 accuracy 0.95 913 macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report: precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	_		0.89	0.90	223	
macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report:	Severe			0.77	51	
macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report:						
macro avg 0.89 0.87 0.88 913 weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report: precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 200 Severe 0.82 0.84 0.83 44 44 accuracy accuracy macro avg 0.91 0.91 0.91 0.91 848 848	accuracy			0.95	913	
weighted avg 0.95 0.95 0.95 913 Autotest dataset - Classification Report: precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	-	0.89	0.87	0.88	913	
Autotest dataset - Classification Report:	_		0.95	0.95	913	
precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	0 0					
precision recall f1-score support Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848						
Healthy 0.98 0.99 0.98 604 Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	Autotest data	set - Class	ification	Report:		
Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848		precision	recall	f1-score	support	
Moderate 0.93 0.90 0.91 200 Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848						
Severe 0.82 0.84 0.83 44 accuracy 0.96 848 macro avg 0.91 0.91 0.91 848	Healthy	0.98	0.99	0.98	604	
accuracy 0.96 848 macro avg 0.91 0.91 848	Moderate	0.93	0.90	0.91	200	
macro avg 0.91 0.91 848	Severe	0.82	0.84	0.83	44	
macro avg 0.91 0.91 848						
macro avg 0.91 0.91 848	accuracy			0.96	848	
•	_	0.91	0.91	0.91	848	
	_	0.96	0.96	0.96	848	
	8					

Round	Global Accuracy
1	71.15 %
2	73.97 %
3	81.13 %
4	82.43 %
5	86.99 %
6	90.46 %
7	86.77 %
8	90.02 %
9	91.32 %
10	91.76 %
11	91.76 %
12	93.06 %



Auto-Test Dataset Confusion Matrix:



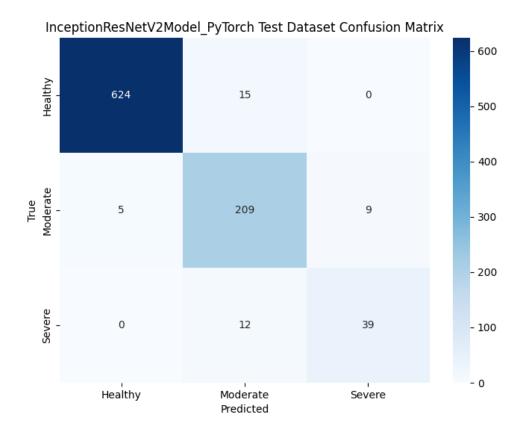
7. InceptionResNetV2 Model – PyTorch Framework – Weight Initialization

Global Model Accuracy: 95.51%

Global Model Loss: 0.1483

	740t <u>L</u> 00	0. 0. 1		
Final Test Results:				
Test Accuracy	. 0 0551			
Test Loss: 0.				
Test Loss. 0.	1403			
AutoTest Resu	ılts:			
AutoTest Accu	racy: 0.962	3		
AutoTest Loss				
Test dataset	- Classific	ation Repo	ort:	
	precision	recall	f1-score	support
Healthy	0.99	0.98	0.98	639
Moderate	0.89	0.94	0.91	223
Severe	0.81	0.76	0.79	51
accuracy			0.96	913
macro avg		0.89	0.89	913
weighted avg	0.96	0.96	0.96	913
Autotest data				
	precision	recall	f1-score	support
n-14h.	0.00	0.00	0.00	504
Healthy			0.98	604
Moderate				200
Severe	0.86	0.84	0.85	44
accuracy			0.96	848
accuracy macro avg	0.92	0.92		848
weighted avg		0.92 0.96		848
wergnred avg	0.96	0.96	0.96	848

Round	Global Accuracy
1	48.81 %
2	87.85 %
3	91.54 %
4	90.02 %
5	92.19 %
6	91.32 %
7	92.62 %
8	94.79 %
9	94.36 %
10	94.14 %
11	94.14 %
12	93.71 %



Auto-Test Dataset Confusion Matrix:

