HIERARCHICAL

==

AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET

==

STAGE 1 (I vs Rest) - AGGREGATED

----
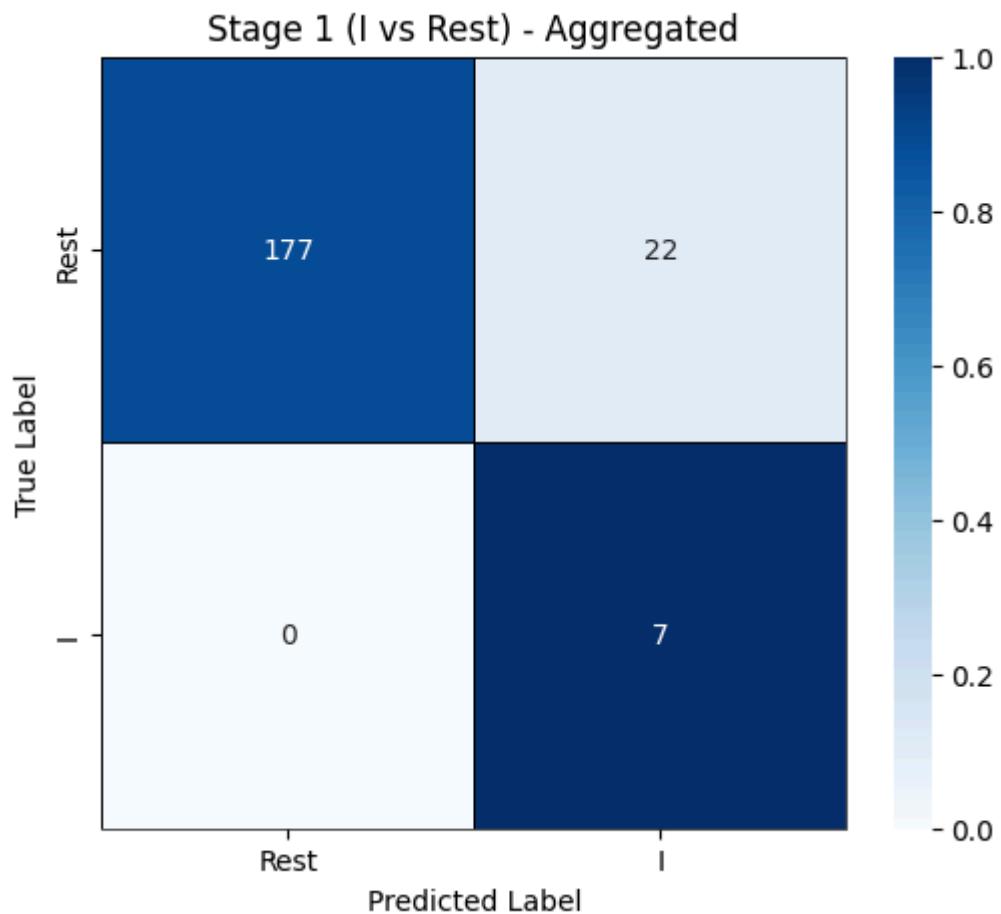
	precision	recall	f1-score	support
0	1.0000	0.8894	0.9415	199
1	0.2414	1.0000	0.3889	7
accuracy			0.8932	206
macro avg	0.6207	0.9447	0.6652	206
weighted avg	0.9742	0.8932	0.9227	206

Balanced Accuracy: 0.9447

Stage 1 (I vs Rest) - Aggregated

STAGE 2 (DC vs Rest) - AGGREGATED

----

	precision	recall	f1-score	support
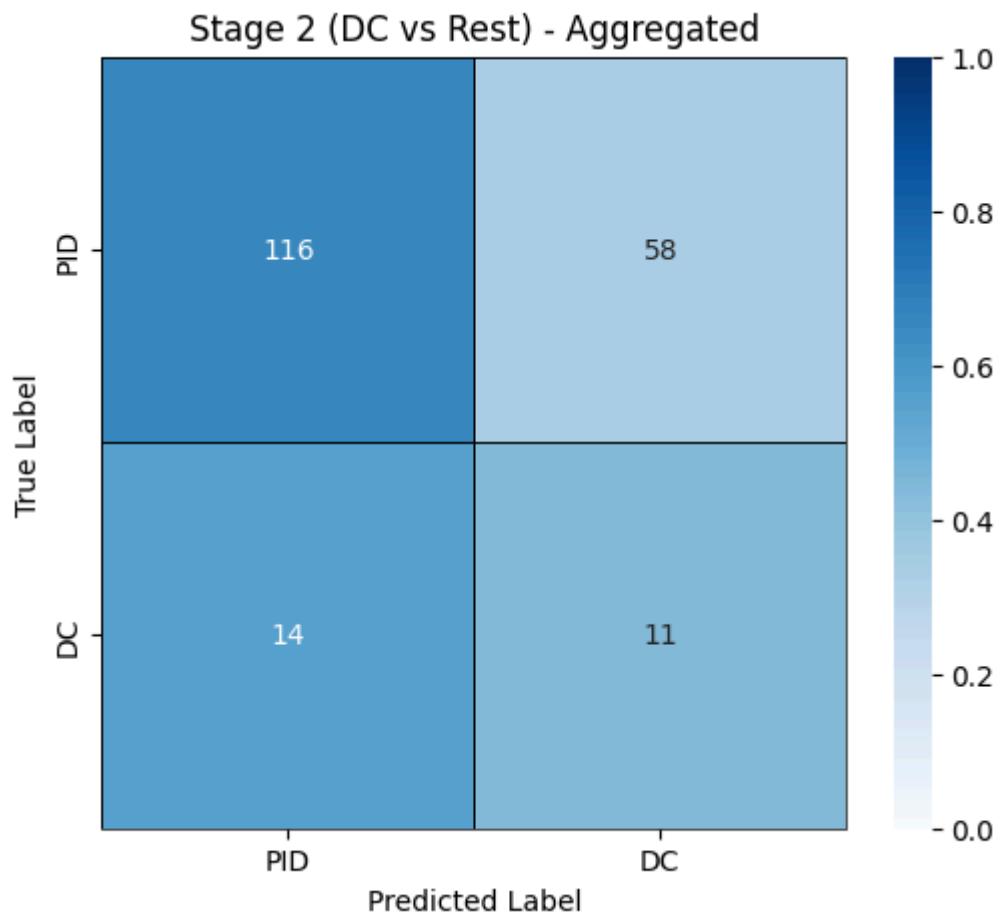
	precision	recall	f1-score	support
0	0.8923	0.6667	0.7632	174
1	0.1594	0.4400	0.2340	25
accuracy			0.6382	199
macro avg	0.5259	0.5533	0.4986	199
weighted avg	0.8002	0.6382	0.6967	199

Balanced Accuracy: 0.5533

Stage 2 (DC vs Rest) - Aggregated
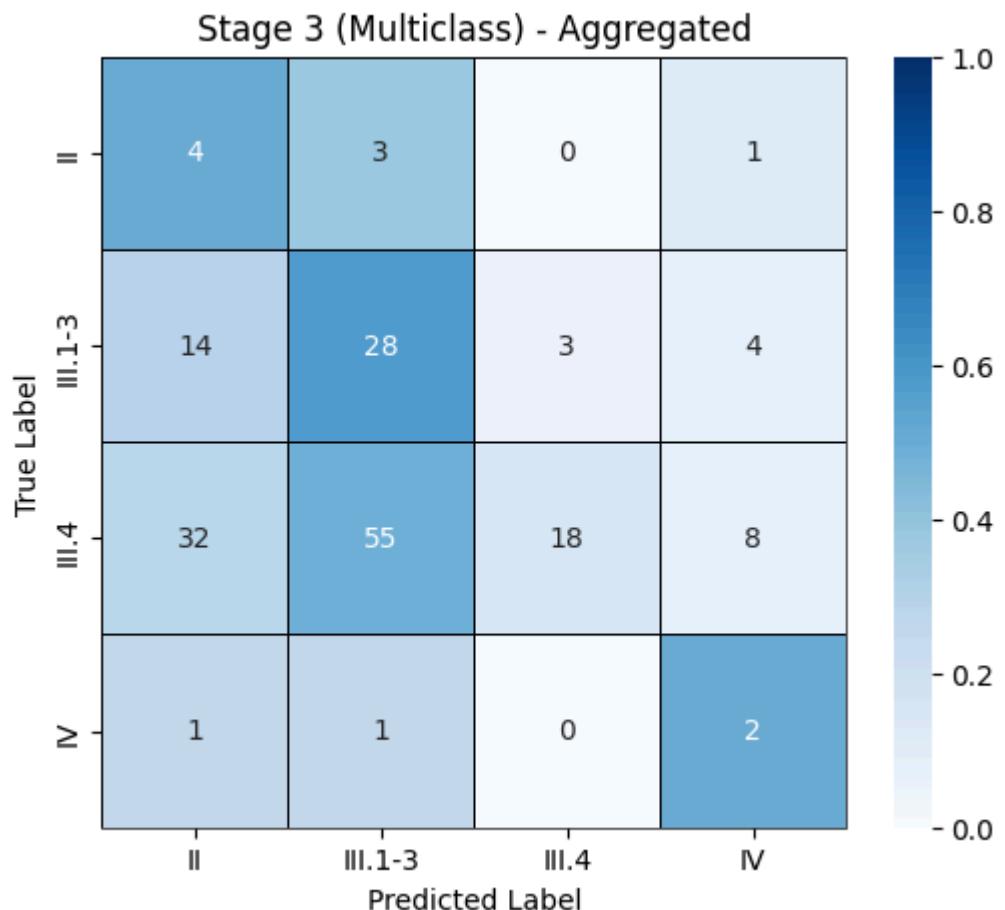
STAGE 3 (Multiclass) - AGGREGATED
----

	precision	recall	f1-score	support
0	0.0784	0.5000	0.1356	8
1	0.3218	0.5714	0.4118	49
2	0.8571	0.1593	0.2687	113
3	0.1333	0.5000	0.2105	4
accuracy			0.2989	174
macro avg	0.3477	0.4327	0.2566	174
weighted avg	0.6540	0.2989	0.3015	174

Balanced Accuracy: 0.4327

Stage 3 (Multiclass) - Aggregated


```
------------------------------------------------------------
SOFT-GATED OVERALL - AGGREGATED
------------------------------------------------------------
```
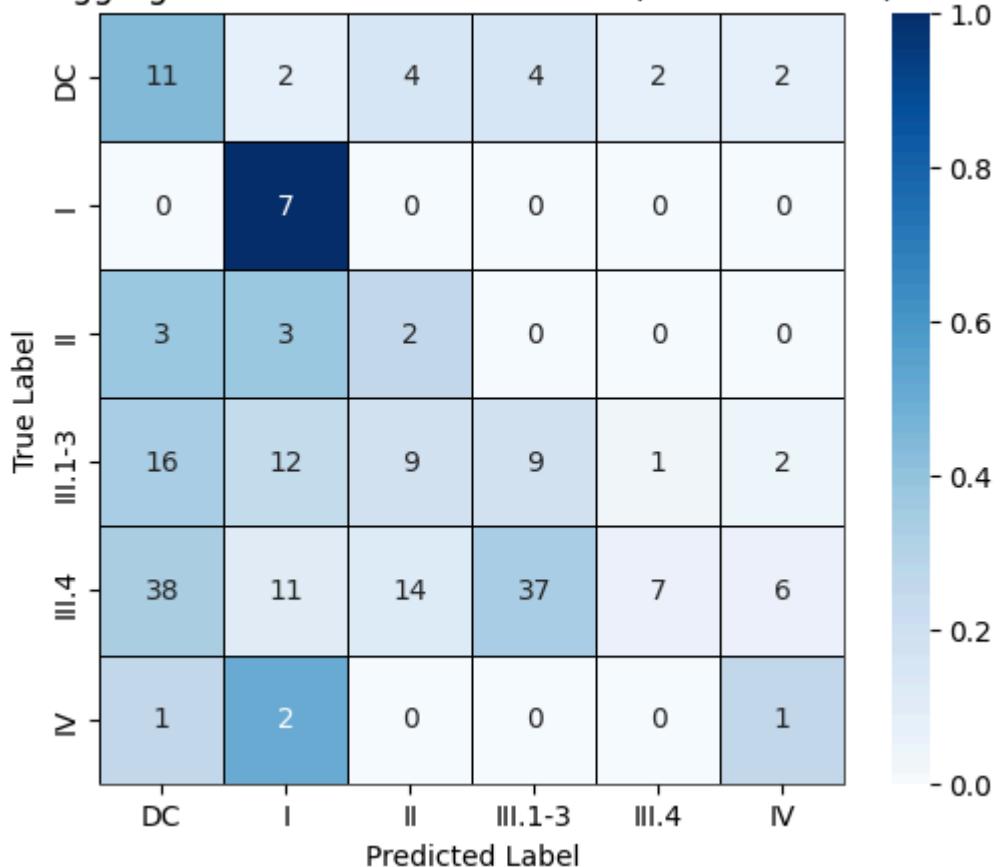
== Aggregated Classification Report ==

	precision	recall	f1-score	support
DC	0.1594	0.4400	0.2340	25
I	0.1892	1.0000	0.3182	7
II	0.0690	0.2500	0.1081	8
III.1-3	0.1800	0.1837	0.1818	49
III.4	0.7000	0.0619	0.1138	113
IV	0.0909	0.2500	0.1333	4
accuracy			0.1796	206
macro avg	0.2314	0.3643	0.1816	206
weighted avg	0.4570	0.1796	0.1517	206

Aggregated Balanced Accuracy: 0.3643

Aggregated Final Confusion Matrix (Entire Holdout)

== Average Stage Balanced Accuracies ==
Stage 1 (I vs Rest): 0.9444
Stage 2 (DC vs Rest): 0.5481
Stage 3 (Multiclass): 0.4318
Final (Soft-Gated): 0.3624

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==
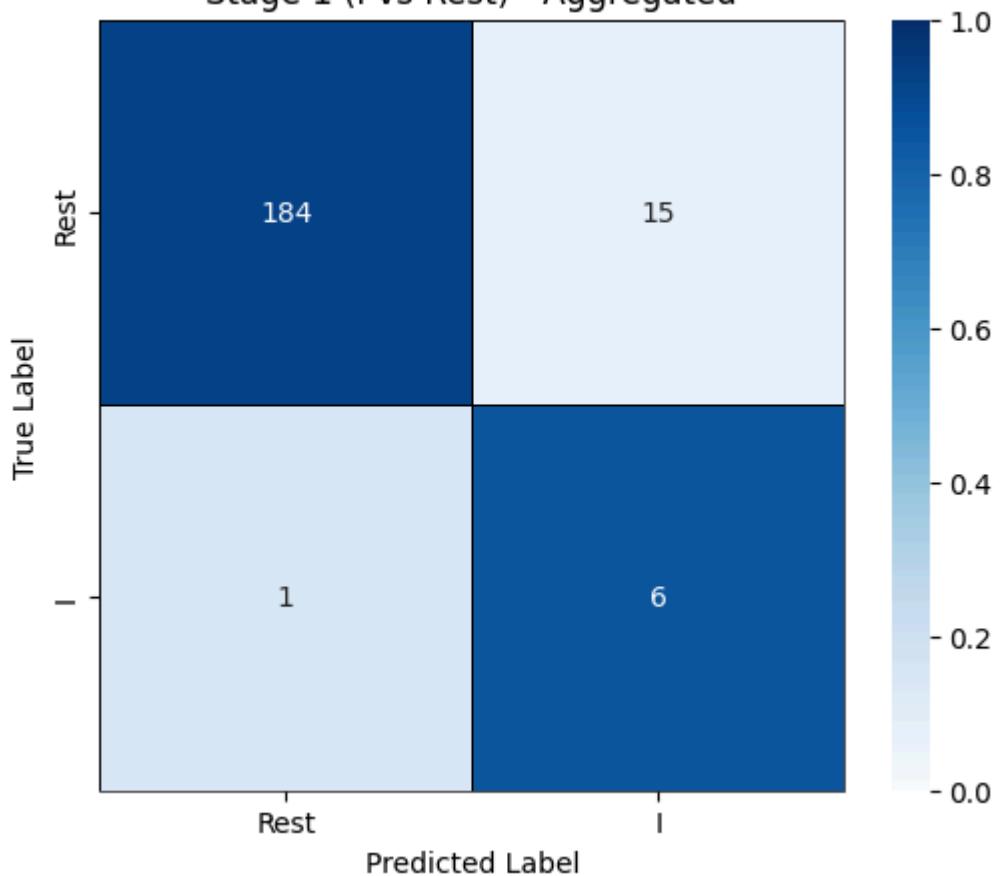
STAGE 1 (I vs Rest) - AGGREGATED
----
          precision   recall f1-score   support

      0   0.9946   0.9246   0.9583   199
      1   0.2857   0.8571   0.4286   7

    accuracy   0.9223   206
   macro avg   0.6402   0.8909   0.6935   206
weighted avg   0.9705   0.9223   0.9403   206

Balanced Accuracy: 0.8909

Stage 1 (I vs Rest) - Aggregated
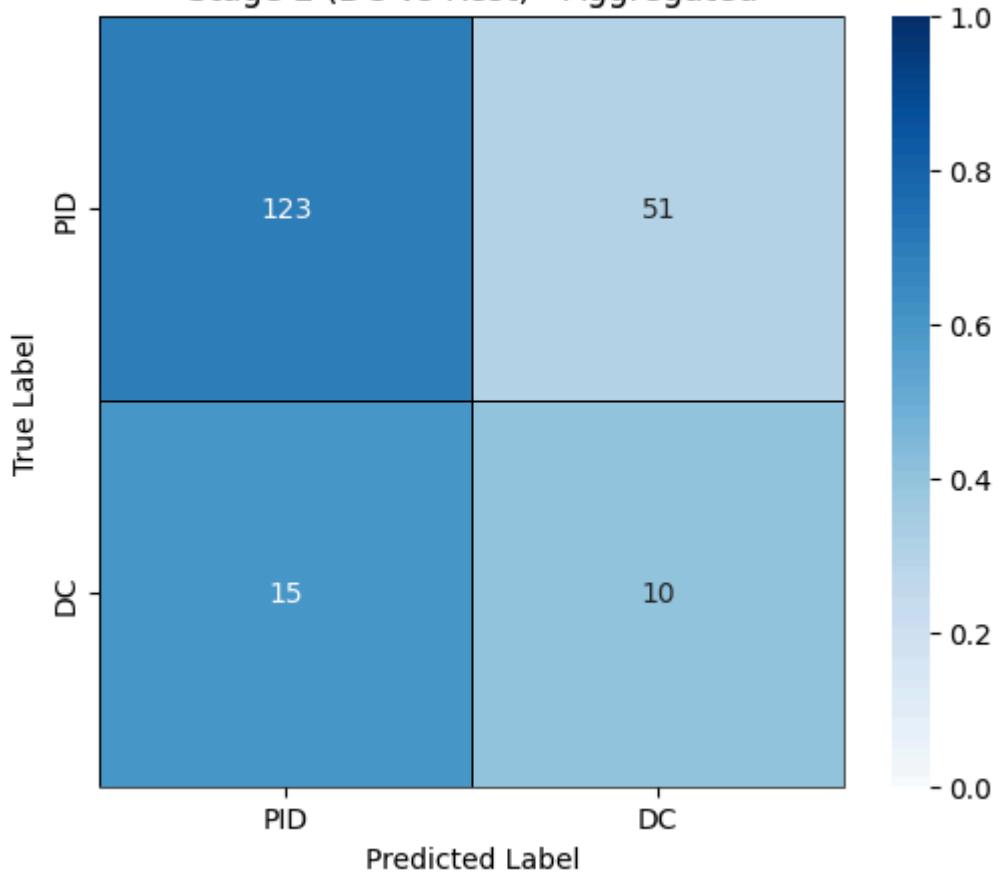

```
--------------------------------------------------------
STAGE 2 (DC vs Rest) - AGGREGATED
--------------------------------------------------------
              precision   recall  f1-score   support

          0    0.8913    0.7069    0.7885       174
          1    0.1639    0.4000    0.2326        25

   accuracy                        0.6683       199
  macro avg    0.5276    0.5534    0.5105       199
weighted avg   0.7999    0.6683    0.7186       199
```

Balanced Accuracy: 0.5534

Stage 2 (DC vs Rest) - Aggregated
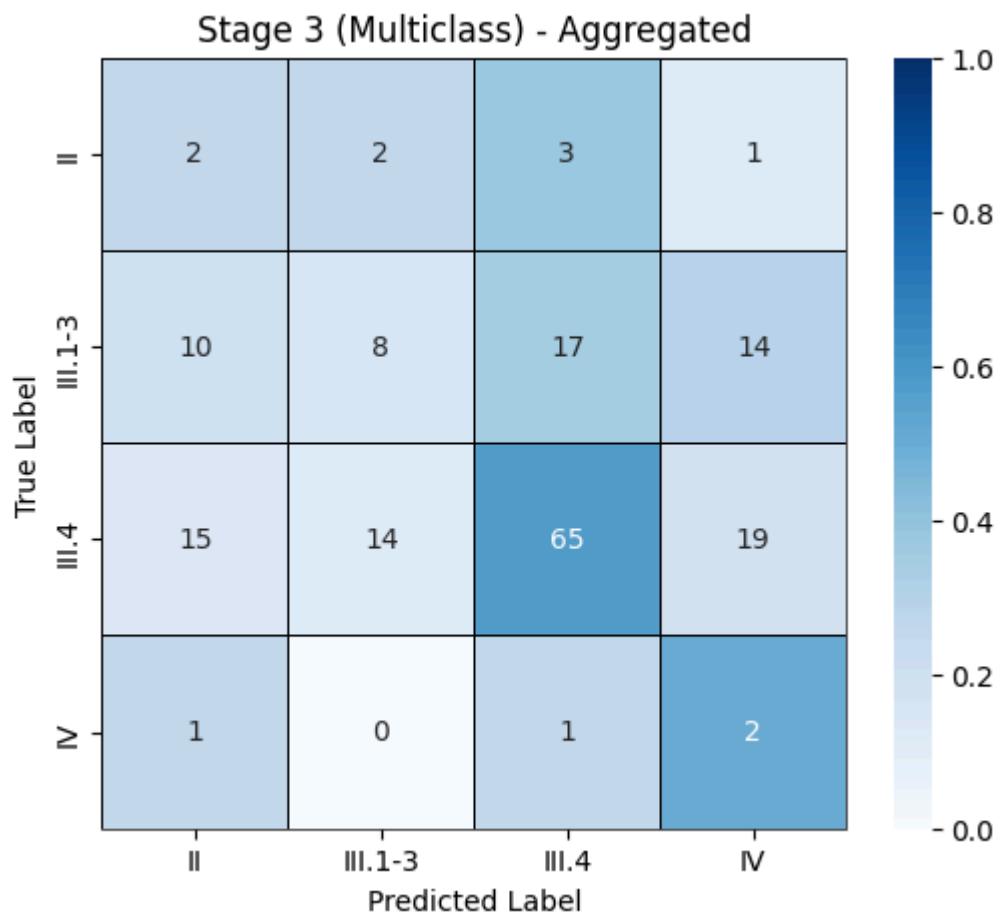
STAGE 3 (Multiclass) - AGGREGATED
----

	precision	recall	f1-score	support
0	0.0714	0.2500	0.1111	8
1	0.3333	0.1633	0.2192	49
2	0.7558	0.5752	0.6533	113
3	0.0556	0.5000	0.1000	4
accuracy			0.4425	174
macro avg	0.3040	0.3721	0.2709	174
weighted avg	0.5893	0.4425	0.4934	174

Balanced Accuracy: 0.3721

Stage 3 (Multiclass) - Aggregated

SOFT-GATED OVERALL - AGGREGATED
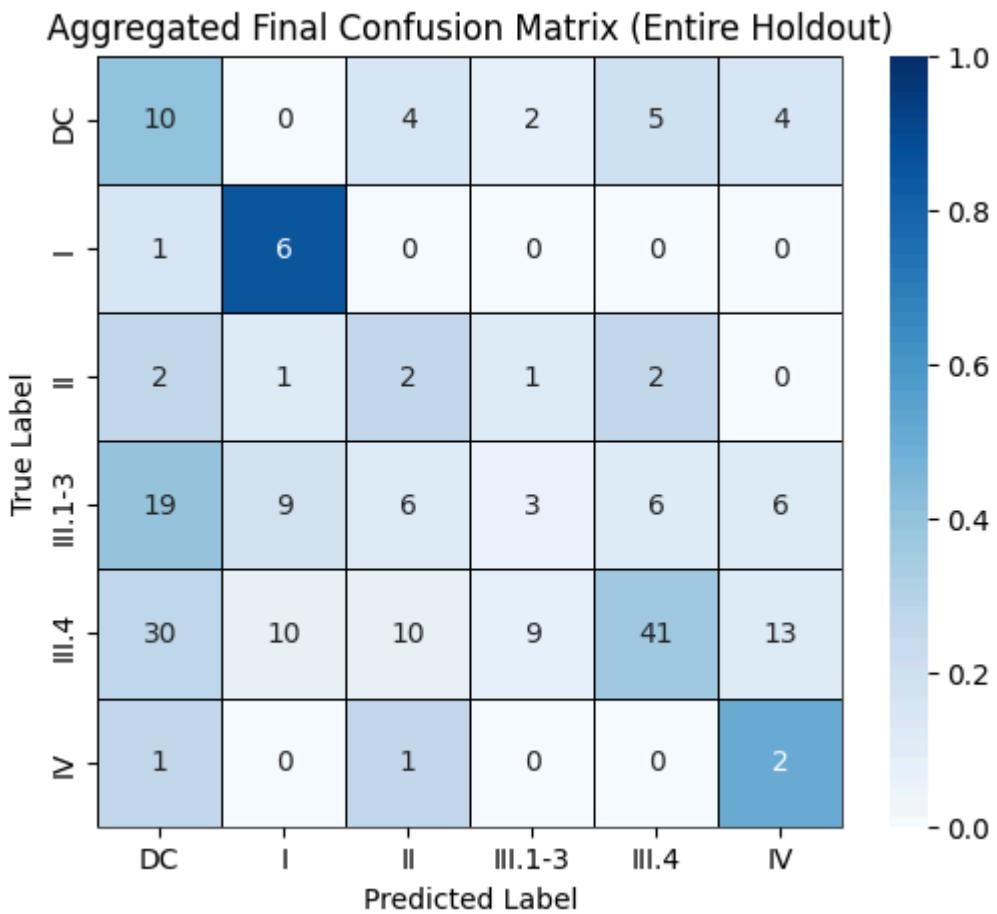
== Aggregated Classification Report ==

	precision	recall	f1-score	support
DC	0.1587	0.4000	0.2273	25
I	0.2308	0.8571	0.3636	7
II	0.0870	0.2500	0.1290	8
III.1-3	0.2000	0.0612	0.0938	49
III.4	0.7593	0.3628	0.4910	113
IV	0.0800	0.5000	0.1379	4
accuracy			0.3107	206
macro avg	0.2526	0.4052	0.2404	206
weighted avg	0.4961	0.3107	0.3393	206

Aggregated Balanced Accuracy: 0.4052

Aggregated Final Confusion Matrix (Entire Holdout)

== Average Stage Balanced Accuracies ==
Stage 1 (I vs Rest): 0.8999
Stage 2 (DC vs Rest): 0.5538
Stage 3 (Multiclass): 0.3724
Final (Soft-Gated): 0.4082

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==
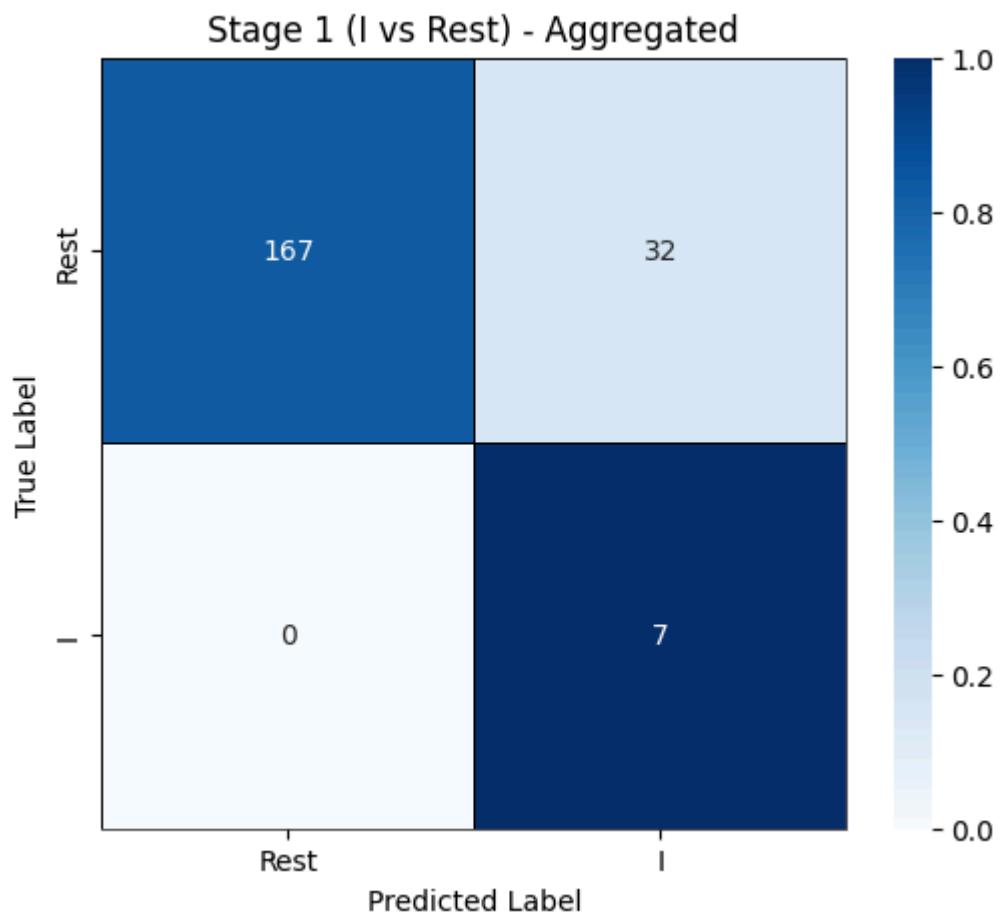
STAGE 1 (I vs Rest) - AGGREGATED
----
         precision   recall f1-score   support

     0   1.0000   0.8392   0.9126   199
     1   0.1795   1.0000   0.3043   7

  accuracy   0.8447   206
 macro avg   0.5897   0.9196   0.6085   206
weighted avg 0.9721   0.8447   0.8919   206

Balanced Accuracy: 0.9196

Stage 1 (I vs Rest) - Aggregated
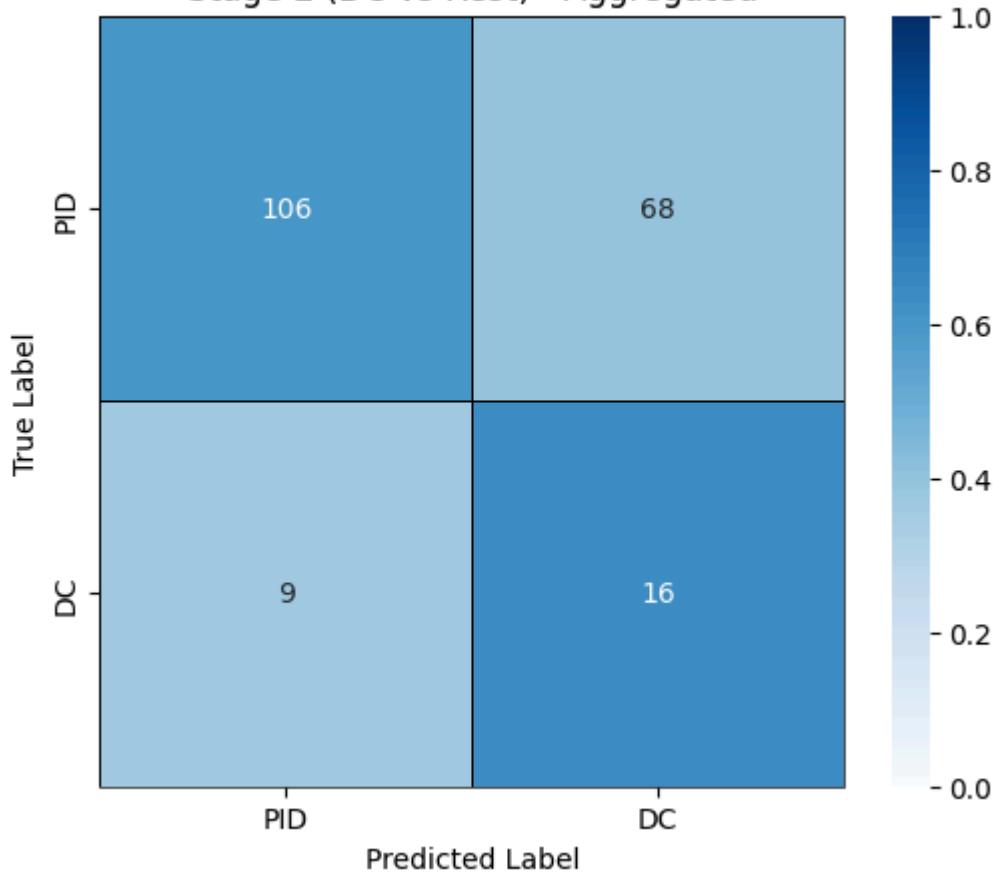

```
----------------------------------------------------
STAGE 2 (DC vs Rest) - AGGREGATED
----------------------------------------------------
              precision   recall  f1-score  support

        0      0.9217     0.6092    0.7336      174
        1      0.1905     0.6400    0.2936       25

  accuracy                         0.6131      199
 macro avg     0.5561     0.6246    0.5136      199
weighted avg   0.8299     0.6131    0.6783      199
```

Balanced Accuracy: 0.6246

Stage 2 (DC vs Rest) - Aggregated

STAGE 3 (Multiclass) - AGGREGATED

----
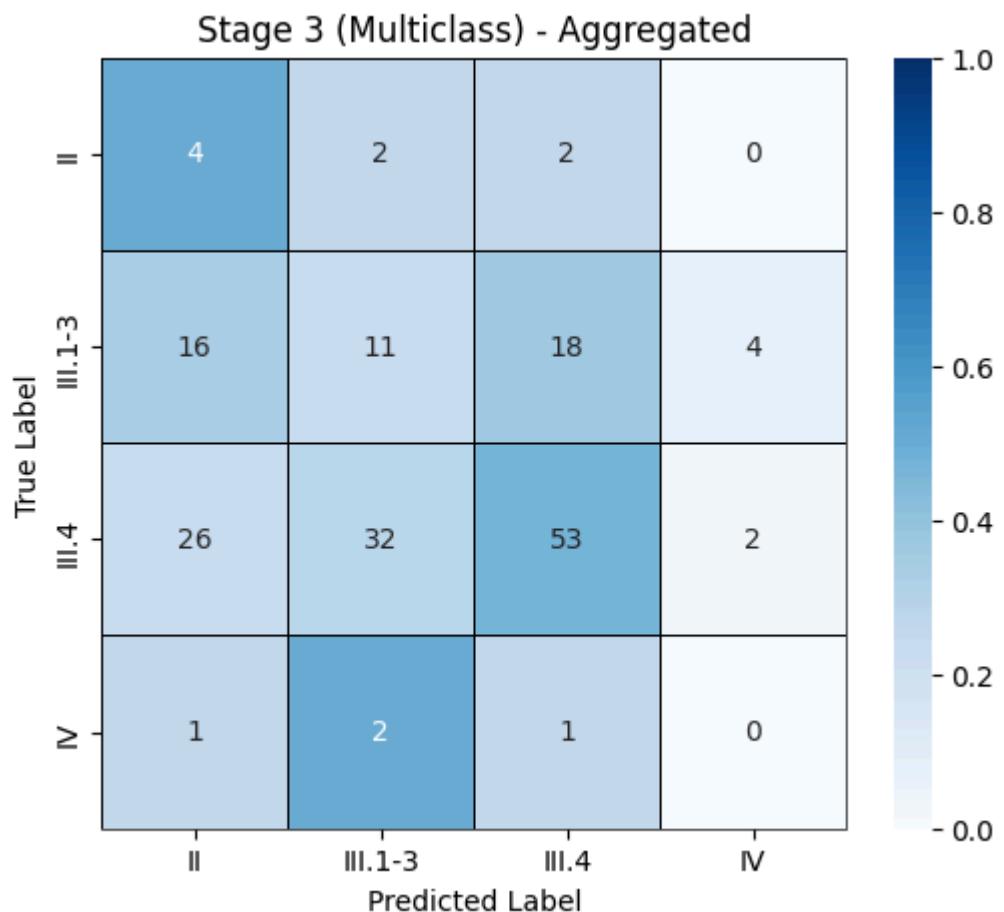
	precision	recall	f1-score	support
0	0.0851	0.5000	0.1455	8
1	0.2340	0.2245	0.2292	49
2	0.7162	0.4690	0.5668	113
3	0.0000	0.0000	0.0000	4
accuracy			0.3908	174
macro avg	0.2588	0.2984	0.2354	174
weighted avg	0.5350	0.3908	0.4393	174

Balanced Accuracy: 0.2984

Stage 3 (Multiclass) - Aggregated

SOFT-GATED OVERALL - AGGREGATED
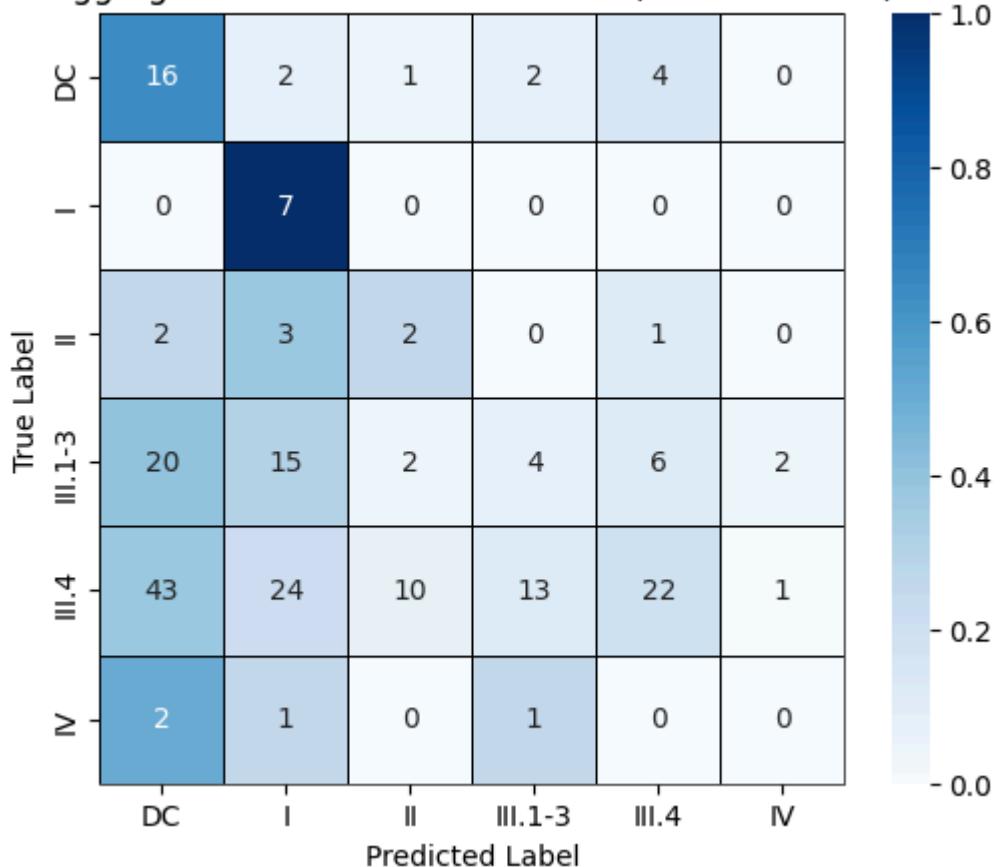
== Aggregated Classification Report ==

	precision	recall	f1-score	support
DC	0.1928	0.6400	0.2963	25
I	0.1346	1.0000	0.2373	7
II	0.1333	0.2500	0.1739	8
III.1-3	0.2000	0.0816	0.1159	49
III.4	0.6667	0.1947	0.3014	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.2476	206
macro avg	0.2212	0.3611	0.1875	206
weighted avg	0.4464	0.2476	0.2437	206

Aggregated Balanced Accuracy: 0.3611

Aggregated Final Confusion Matrix (Entire Holdout)

== Average Stage Balanced Accuracies ==
Stage 1 (I vs Rest): 0.9196
Stage 2 (DC vs Rest): 0.6203
Stage 3 (Multiclass): 0.2982
Final (Soft-Gated): 0.3597

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==

STAGE 1 (I vs Rest) - AGGREGATED
----
           precision   recall f1-score   support

        0   0.9946   0.9246   0.9583   199
        1   0.2857   0.8571   0.4286   7
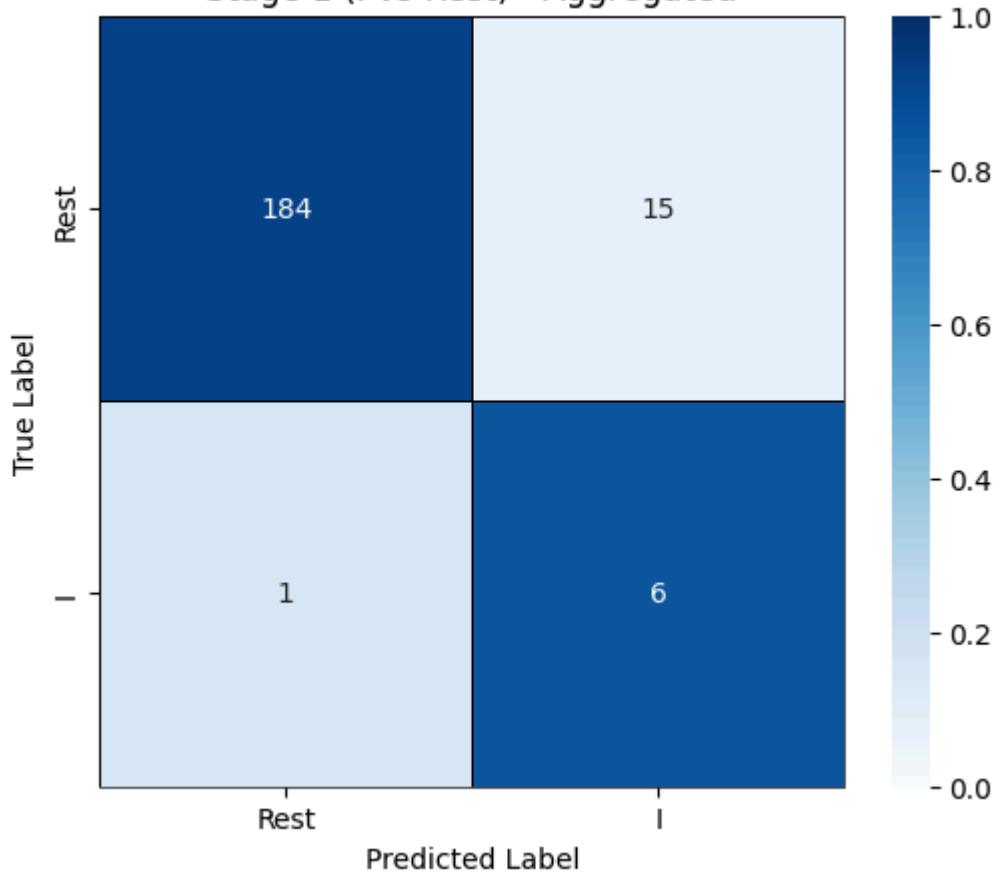
 accuracy   0.9223   206
 macro avg   0.6402   0.8909   0.6935   206
weighted avg   0.9705   0.9223   0.9403   206

Balanced Accuracy: 0.8909

Stage 1 (I vs Rest) - Aggregated
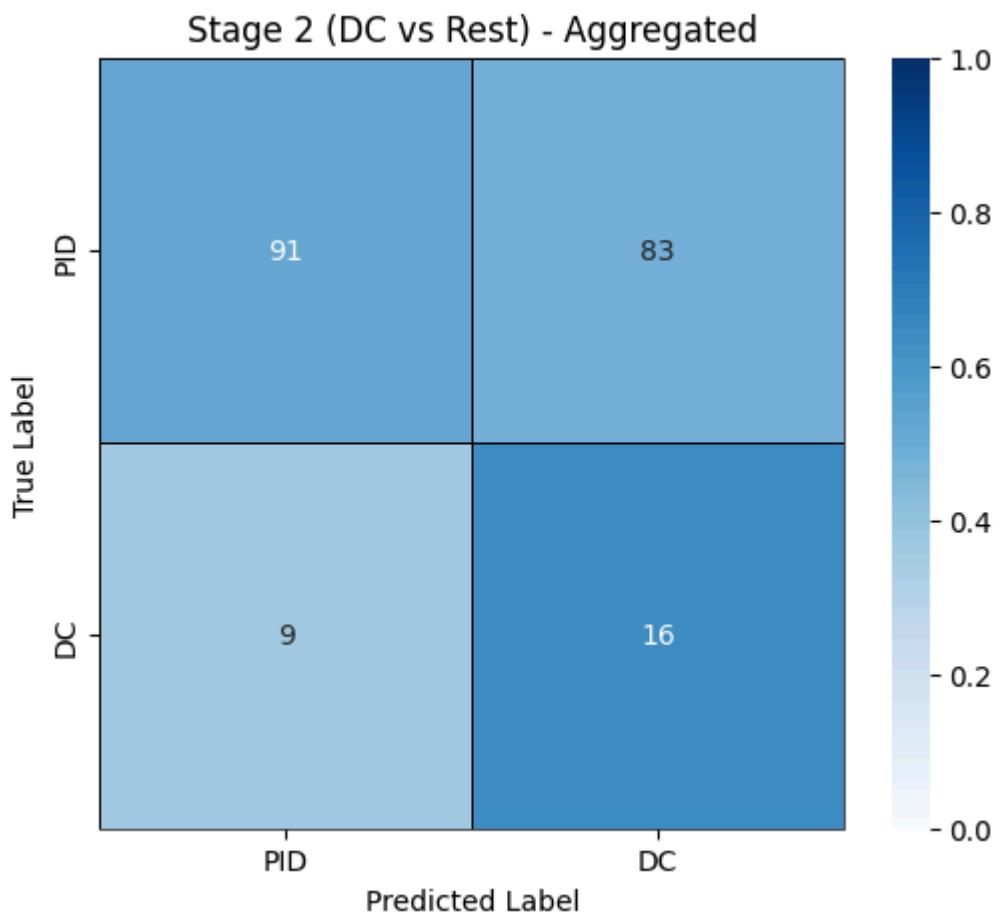

```
------------------------------------------------------
STAGE 2 (DC vs Rest) - AGGREGATED
------------------------------------------------------
              precision   recall  f1-score   support

         0      0.9100    0.5230    0.6642       174
         1      0.1616    0.6400    0.2581        25

  accuracy                          0.5377       199
 macro avg      0.5358    0.5815    0.4611       199
weighted avg    0.8160    0.5377    0.6132       199

Balanced Accuracy: 0.5815
```

Stage 2 (DC vs Rest) - Aggregated
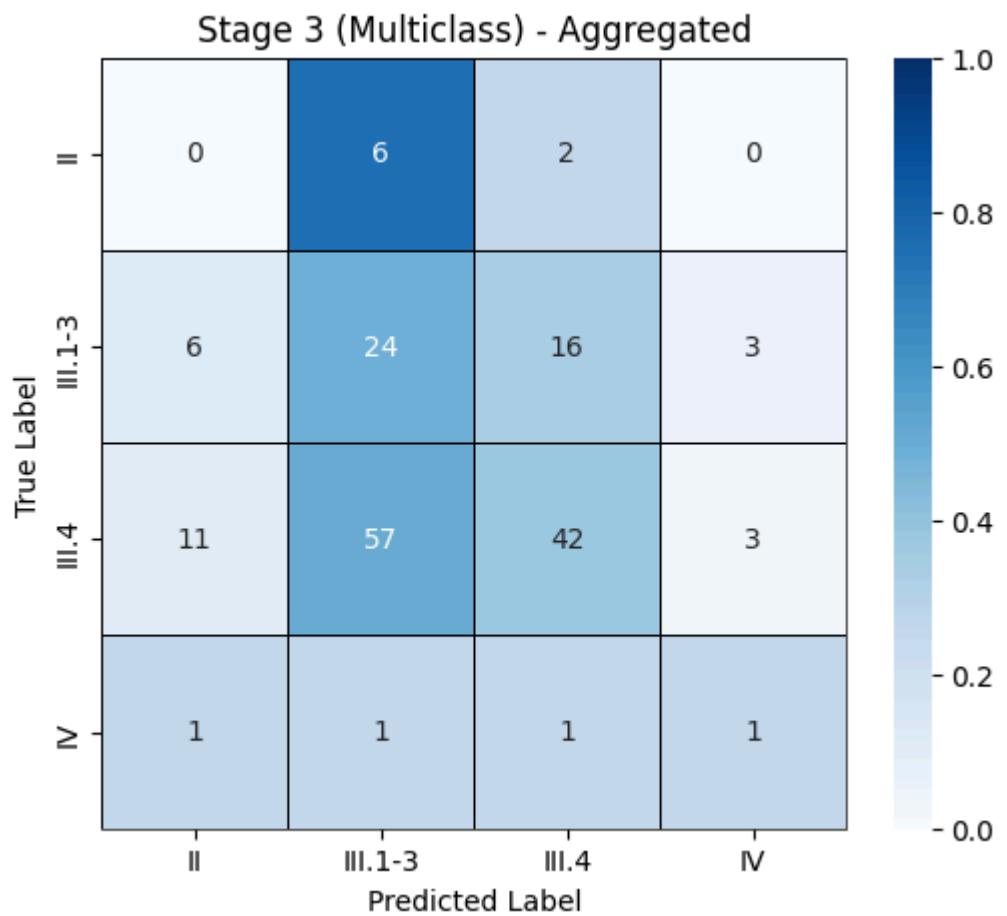
STAGE 3 (Multiclass) - AGGREGATED
----

	precision	recall	f1-score	support
0	0.0000	0.0000	0.0000	8
1	0.2727	0.4898	0.3504	49
2	0.6885	0.3717	0.4828	113
3	0.1429	0.2500	0.1818	4
accuracy			0.3851	174
macro avg	0.2760	0.2779	0.2537	174
weighted avg	0.5272	0.3851	0.4164	174

Balanced Accuracy: 0.2779

Stage 3 (Multiclass) - Aggregated

SOFT-GATED OVERALL - AGGREGATED
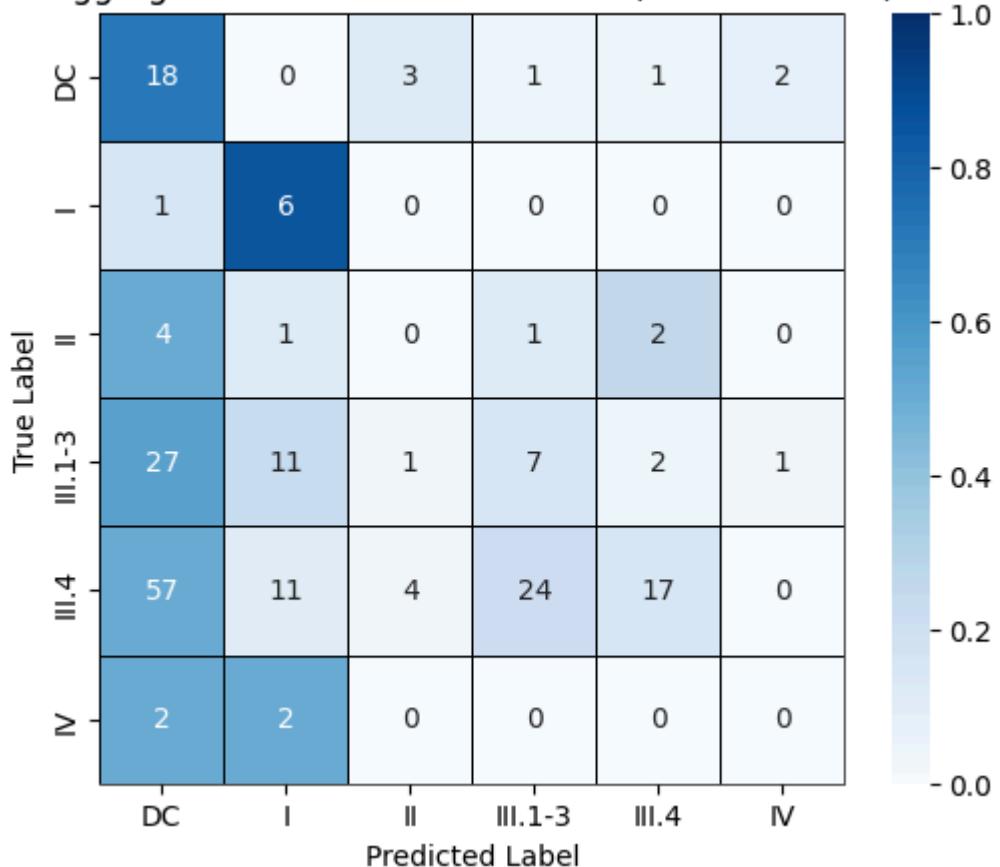
== Aggregated Classification Report ==

	precision	recall	f1-score	support
DC	0.1651	0.7200	0.2687	25
I	0.1935	0.8571	0.3158	7
II	0.0000	0.0000	0.0000	8
III.1-3	0.2121	0.1429	0.1707	49
III.4	0.7727	0.1504	0.2519	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.2330	206
macro avg	0.2239	0.3117	0.1678	206
weighted avg	0.5009	0.2330	0.2221	206

Aggregated Balanced Accuracy: 0.3117

Aggregated Final Confusion Matrix (Entire Holdout)

== Average Stage Balanced Accuracies ==
Stage 1 (I vs Rest): 0.8997
Stage 2 (DC vs Rest): 0.5756
Stage 3 (Multiclass): 0.2762
Final (Soft-Gated): 0.3127

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==
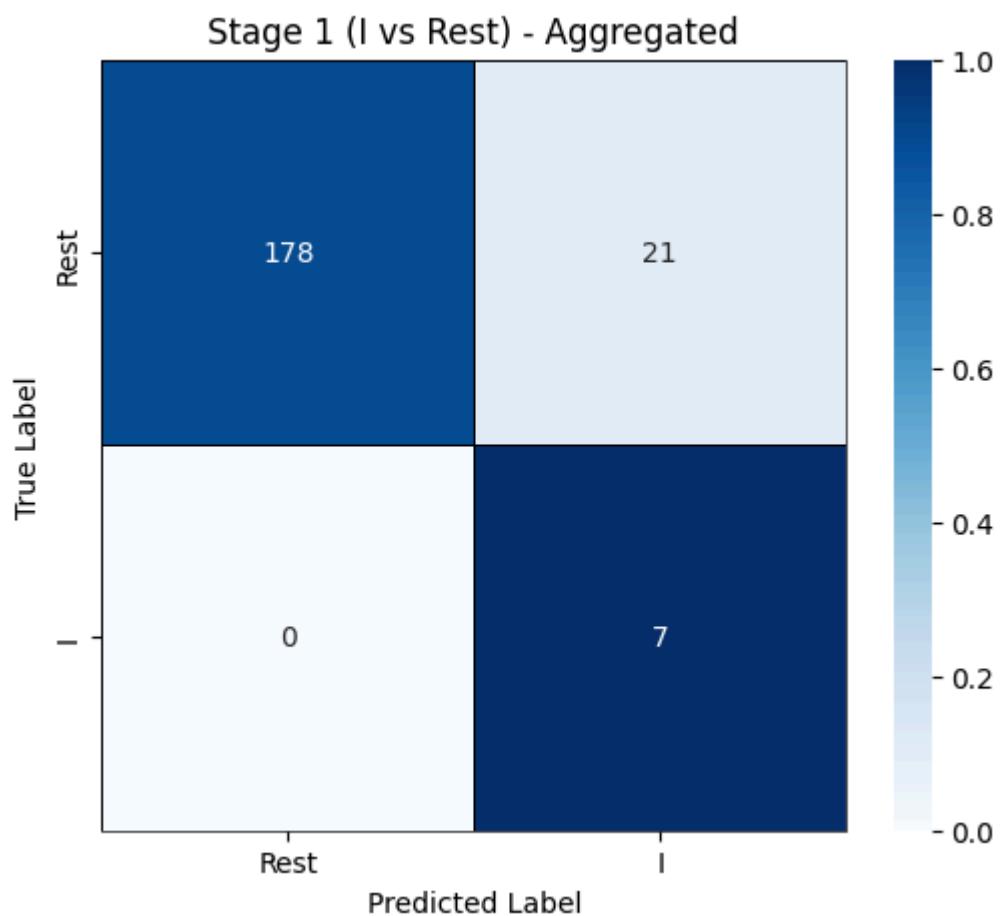
STAGE 1 (I vs Rest) - AGGREGATED
----
          precision   recall f1-score   support

      0   1.0000   0.8945   0.9443   199
      1   0.2500   1.0000   0.4000   7

   accuracy   0.8981   206
 macro avg   0.6250   0.9472   0.6721   206
weighted avg   0.9745   0.8981   0.9258   206

Balanced Accuracy: 0.9472

Stage 1 (I vs Rest) - Aggregated

STAGE 2 (DC vs Rest) - AGGREGATED

----
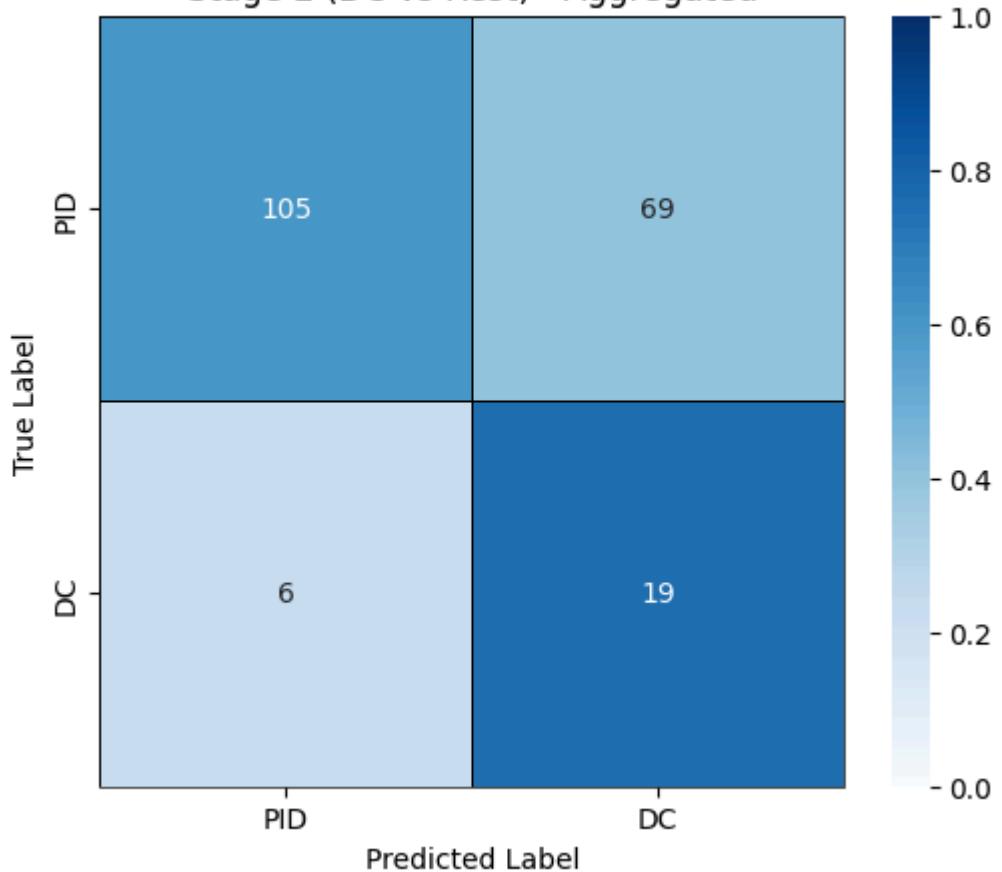
	precision	recall	f1-score	support
0	0.9459	0.6034	0.7368	174
1	0.2159	0.7600	0.3363	25
accuracy			0.6231	199
macro avg	0.5809	0.6817	0.5366	199
weighted avg	0.8542	0.6231	0.6865	199

Balanced Accuracy: 0.6817

Stage 2 (DC vs Rest) - Aggregated

```
-------------------------------------------------------
STAGE 3 (Multiclass) - AGGREGATED
-------------------------------------------------------
```
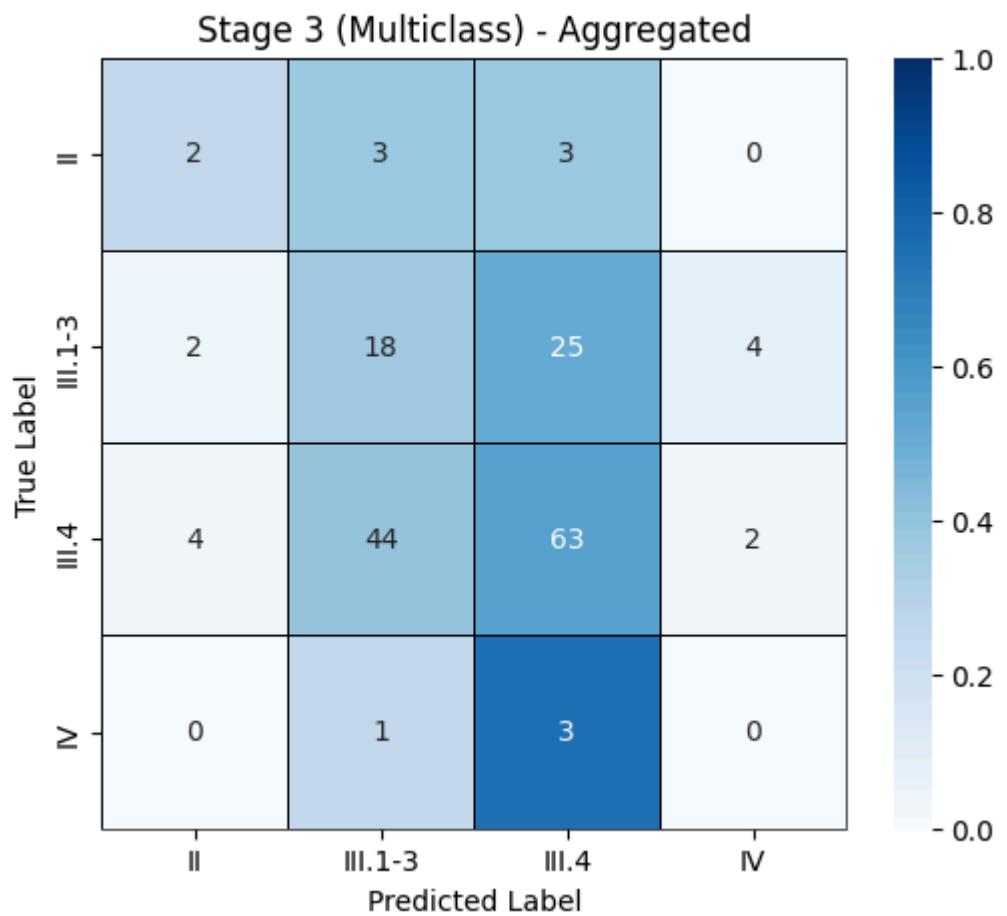
	precision	recall	f1-score	support
0	0.2500	0.2500	0.2500	8
1	0.2727	0.3673	0.3130	49
2	0.6702	0.5575	0.6087	113
3	0.0000	0.0000	0.0000	4
accuracy			0.4770	174
macro avg	0.2982	0.2937	0.2929	174
weighted avg	0.5235	0.4770	0.4950	174

Balanced Accuracy: 0.2937

Stage 3 (Multiclass) - Aggregated

SOFT-GATED OVERALL - AGGREGATED
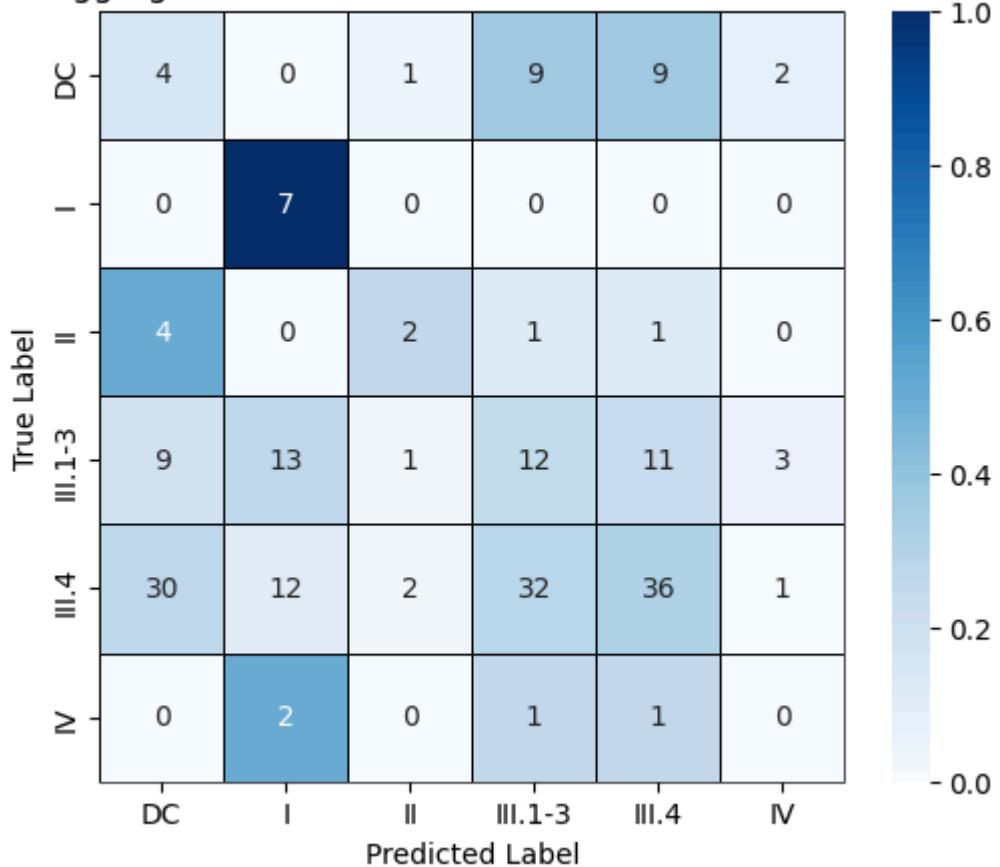
== Aggregated Classification Report ==

	precision	recall	f1-score	support
DC	0.0851	0.1600	0.1111	25
I	0.2059	1.0000	0.3415	7
II	0.3333	0.2500	0.2857	8
III.1-3	0.2182	0.2449	0.2308	49
III.4	0.6207	0.3186	0.4211	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.2961	206
macro avg	0.2439	0.3289	0.2317	206
weighted avg	0.4226	0.2961	0.3220	206

Aggregated Balanced Accuracy: 0.3289

Aggregated Final Confusion Matrix (Entire Holdout)

== Average Stage Balanced Accuracies ==
Stage 1 (I vs Rest): 0.9472
Stage 2 (DC vs Rest): 0.4668
Stage 3 (Multiclass): 0.2951
Final (Soft-Gated): 0.3302

[0.9472361809045227, 0.6817241379310345, 0.29371726566732886, 0.3289136716633556]
[0.9447236180904522, 0.5533333333333333, 0.4326801517067004, 0.36427006200710377]
[0.9195979899497487, 0.6245977011494253, 0.29837908614773345, 0.3610538197579917]
[0.8908829863603733, 0.5534482758620689, 0.3721216362651255, 0.40519986755764253]
[0.8908829863603733, 0.5814942528735632, 0.2778693335741376, 0.311740412979351]

0.919 ± 0.011
0.599 ± 0.022
0.335 ± 0.026
0.354 ± 0.014

FLAT

==

RUN 1: Using Split A for validation, Split B for testing

==

These features will be dropped:

['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ──────────────────────────────── **0s** 35ms/step

== Test Set Evaluation - Run 1 ==

	precision	recall	f1-score	support
DC	0.6667	0.1538	0.2500	13
I	0.3333	0.6667	0.4444	3
II	0.2000	0.5000	0.2857	4
IIIa	0.2000	0.5833	0.2979	24
IIIb	0.5000	0.1228	0.1972	57
IV	0.0000	0.0000	0.0000	2
accuracy			0.2621	103
macro avg	0.3167	0.3378	0.2459	103
weighted avg	0.4249	0.2621	0.2341	103

Balanced Accuracy: 0.3377755285650022

Confusion Matrix:

```
[[ 2  0  0  7  4  0]
 [ 0  2  1  0  0  0]
 [ 0  0  2  2  0  0]
 [ 0  3  4 14  3  0]
 [ 1  1  3 45  7  0]
 [ 0  0  0  2  0  0]]
```

Test Confusion Matrix - Run 1


```
============================================================
RUN 2: Using Split B for validation, Split A for testing (SWAP)
============================================================
```

These features will be dropped:
['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

1/4 ━━━━━━━━━━━━━━━━━━━━━━ **0s** 71ms/step
4/4 ━━━━━━━━━━━━━━━━━━━━━━ **0s** 22ms/step

== Test Set Evaluation - Run 2 ==
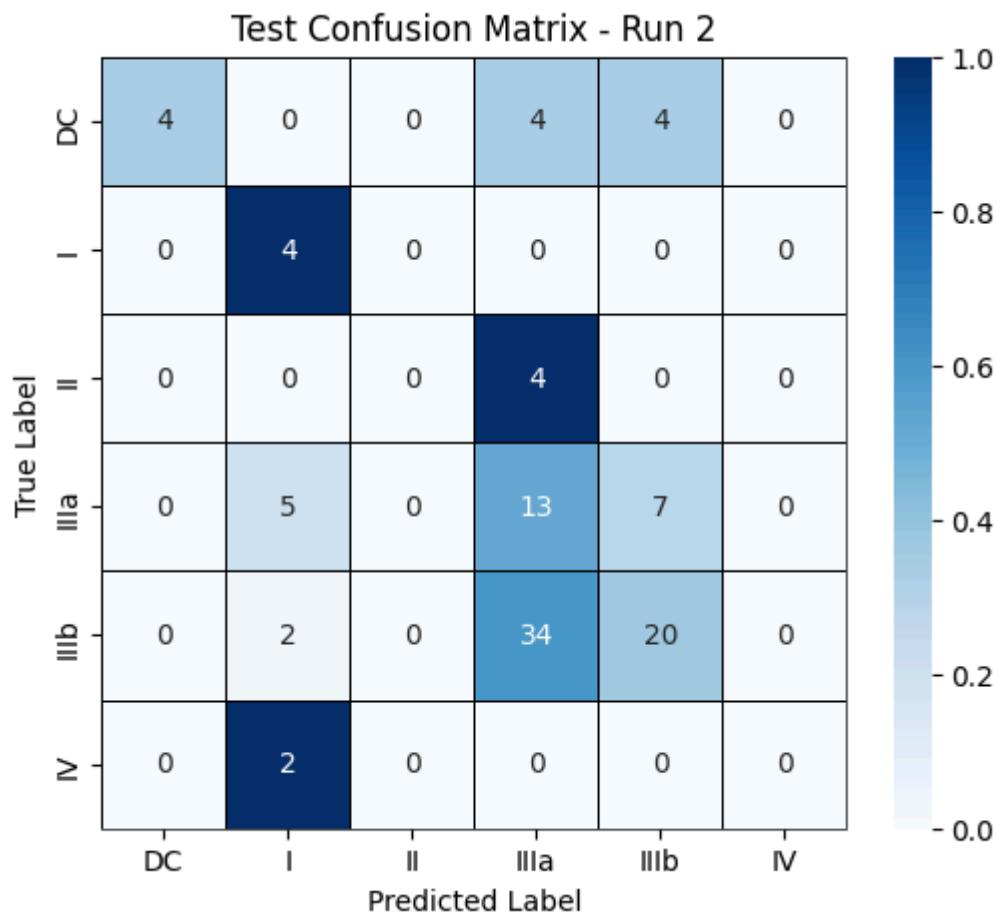
	precision	recall	f1-score	support
DC	1.0000	0.3333	0.5000	12
I	0.3077	1.0000	0.4706	4
II	0.0000	0.0000	0.0000	4
IIIa	0.2364	0.5200	0.3250	25
IIIb	0.6452	0.3571	0.4598	56
IV	0.0000	0.0000	0.0000	2
accuracy			0.3981	103
macro avg	0.3649	0.3684	0.2926	103
weighted avg	0.5366	0.3981	0.4054	103

Balanced Accuracy: 0.36841269841269836
Confusion Matrix:
 [[4 0 0 4 4 0]
 [0 4 0 0 0 0]
 [0 0 0 4 0 0]
 [0 5 0 13 7 0]
 [0 2 0 34 20 0]
 [0 2 0 0 0 0]]

Test Confusion Matrix - Run 2

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==

== Aggregated Classification Report ==
        precision   recall f1-score   support
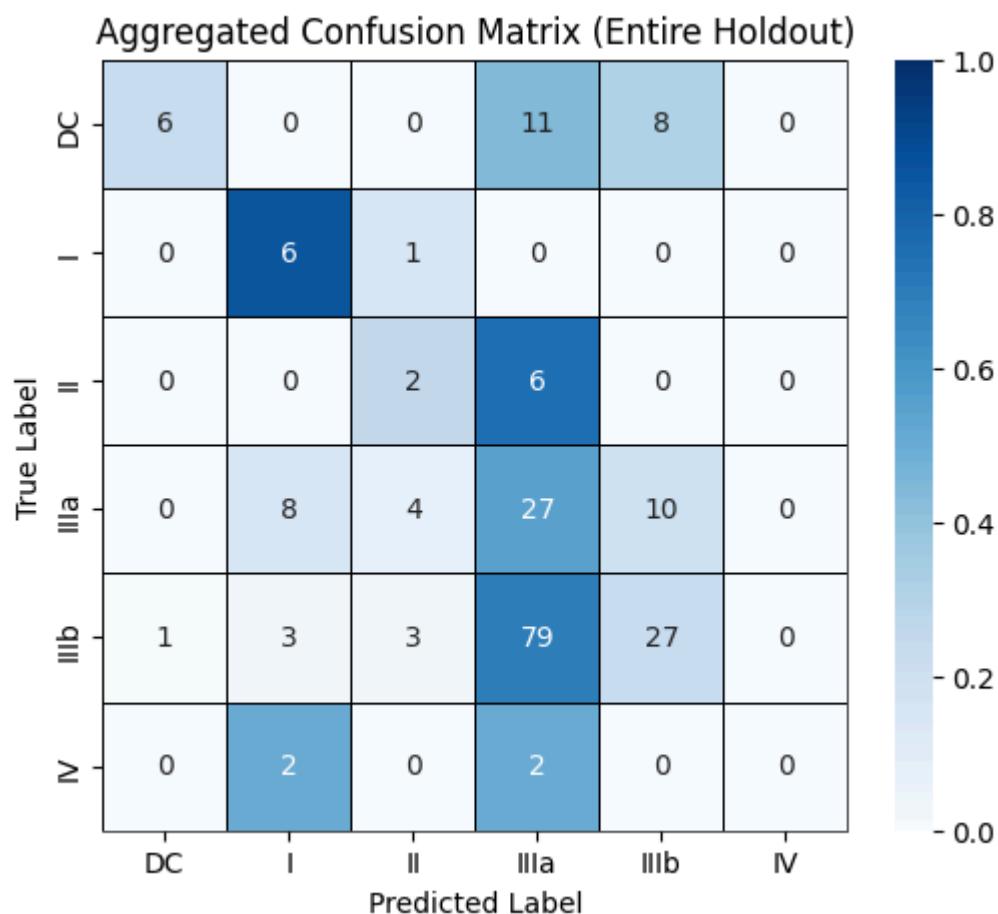
    DC   0.8571   0.2400   0.3750   25
   I   0.3158   0.8571   0.4615   7
    II   0.2000   0.2500   0.2222   8
   IIIa   0.2160   0.5510   0.3103   49
   IIIb   0.6000   0.2389   0.3418   113
    IV   0.0000   0.0000   0.0000   4

```
    accuracy                 0.3301    206
   macro avg    0.3648   0.3562   0.2851      206
weighted avg    0.5030   0.3301   0.3311      206
```

Aggregated Balanced Accuracy: 0.3562
Aggregated Confusion Matrix:
```
[[ 6  0  0 11  8  0]
 [ 0  6  1  0  0  0]
 [ 0  0  2  6  0  0]
 [ 0  8  4 27 10  0]
 [ 1  3  3 79 27  0]
 [ 0  2  0  2  0  0]]
```


Aggregated Confusion Matrix (Entire Holdout)

== Average Balanced Accuracy Across Runs ==
Run 1: 0.3378
Run 2: 0.3684
Average: 0.3531

==
RUN 1: Using Split A for validation, Split B for testing
==
These features will be dropped:

['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ────────────────────────── **0s** 26ms/step

== Test Set Evaluation - Run 1 ==
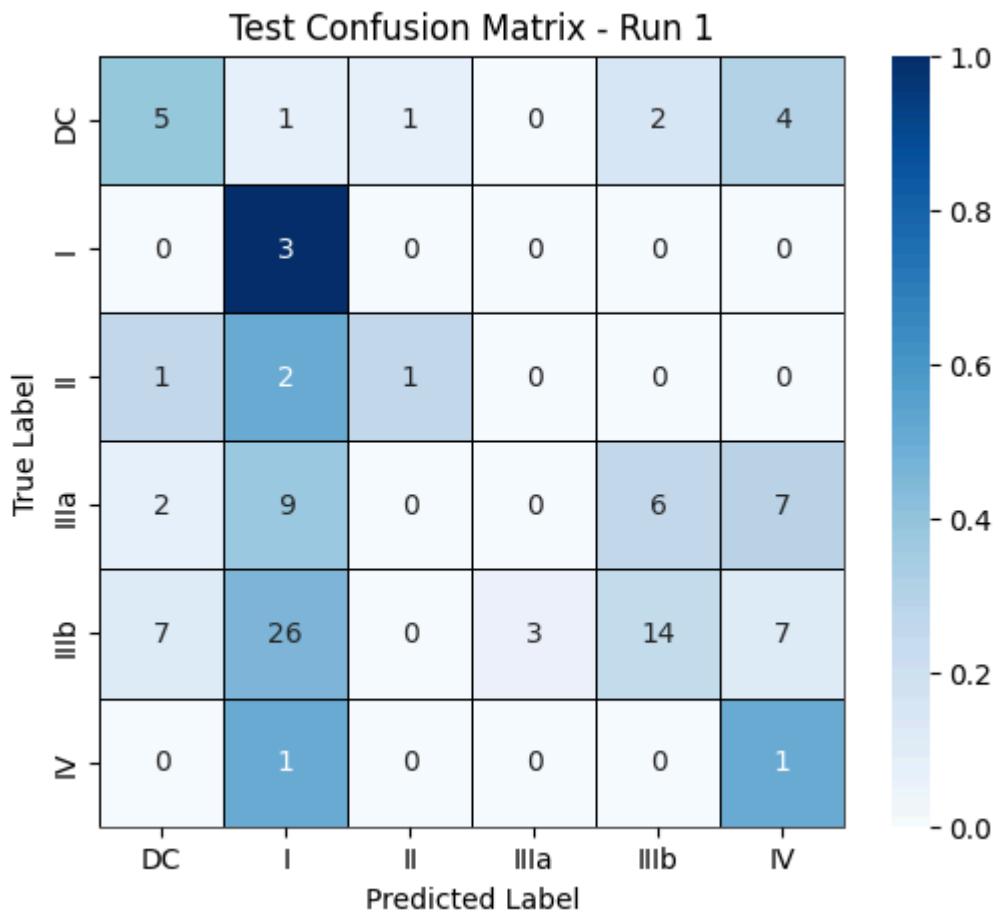
	precision	recall	f1-score	support
DC	0.3333	0.3846	0.3571	13
I	0.0714	1.0000	0.1333	3
II	0.5000	0.2500	0.3333	4
IIIa	0.0000	0.0000	0.0000	24
IIIb	0.6364	0.2456	0.3544	57
IV	0.0526	0.5000	0.0952	2
accuracy			0.2330	103
macro avg	0.2656	0.3967	0.2122	103
weighted avg	0.4168	0.2330	0.2599	103

Balanced Accuracy: 0.3967049032838506
Confusion Matrix:
[[5 1 1 0 2 4]
 [0 3 0 0 0 0]
 [1 2 1 0 0 0]
 [2 9 0 0 6 7]
 [7 26 0 3 14 7]
 [0 1 0 0 0 1]]

Test Confusion Matrix - Run 1

==
RUN 2: Using Split B for validation, Split A for testing (SWAP)
==
These features will be dropped:
['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ───────────────────────────── **0s** 21ms/step

== Test Set Evaluation - Run 2 ==
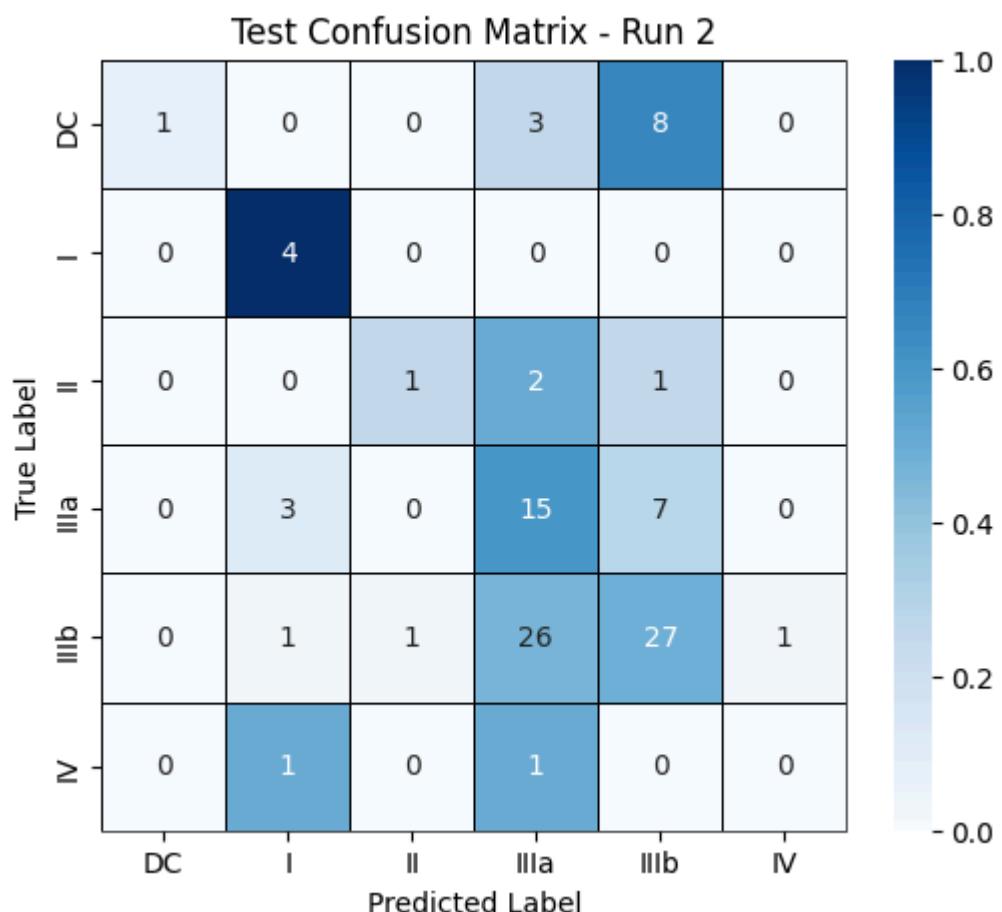
	precision	recall	f1-score	support
DC	1.0000	0.0833	0.1538	12
I	0.4444	1.0000	0.6154	4
II	0.5000	0.2500	0.3333	4
IIIa	0.3191	0.6000	0.4167	25
IIIb	0.6279	0.4821	0.5455	56
IV	0.0000	0.0000	0.0000	2
accuracy			0.4660	103
macro avg	0.4819	0.4026	0.3441	103
weighted avg	0.5720	0.4660	0.4525	103

Balanced Accuracy: 0.40257936507936504

Confusion Matrix:
[[1 0 0 3 8 0]
 [0 4 0 0 0 0]
 [0 0 1 2 1 0]
 [0 3 0 15 7 0]
 [0 1 1 26 27 1]
 [0 1 0 1 0 0]]

Test Confusion Matrix - Run 2

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==

== Aggregated Classification Report ==

|	precision	recall	f1-score	support
DC	0.3750	0.2400	0.2927	25
I	0.1373	1.0000	0.2414	7
II	0.5000	0.2500	0.3333	8
IIIa	0.3000	0.3061	0.3030	49
IIIb	0.6308	0.3628	0.4607	113
IV	0.0500	0.2500	0.0833	4
accuracy			0.3495	206

```
     macro avg     0.3322   0.4015   0.2857      206
weighted avg     0.4879   0.3495   0.3831      206
```

Aggregated Balanced Accuracy: 0.4015
Aggregated Confusion Matrix:
[[6 1 1 3 10 4]
 [0 7 0 0 0 0]
 [1 2 2 2 1 0]
 [2 12 0 15 13 7]
 [7 27 1 29 41 8]
 [0 2 0 1 0 1]]
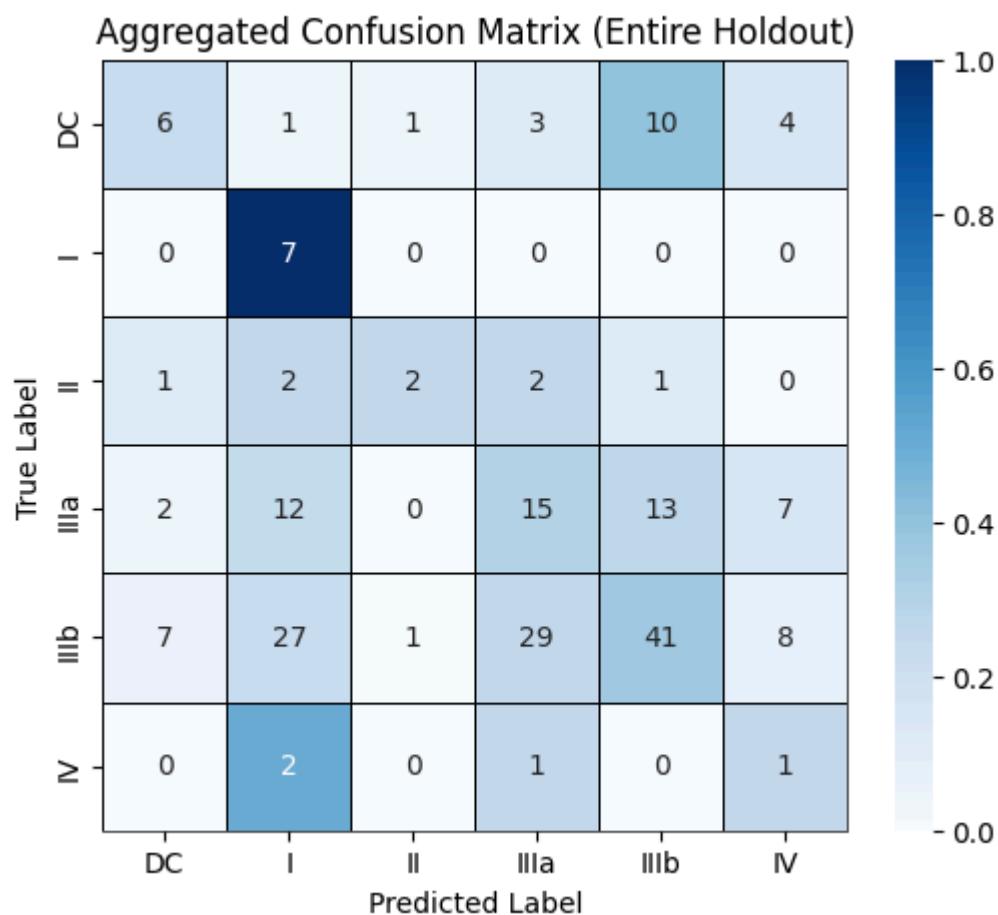
Aggregated Confusion Matrix (Entire Holdout)

== Average Balanced Accuracy Across Runs ==
Run 1: 0.3967
Run 2: 0.4026
Average: 0.3996

==
RUN 1: Using Split A for validation, Split B for testing
==
These features will be dropped:

['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

== Test Set Evaluation - Run 1 ==
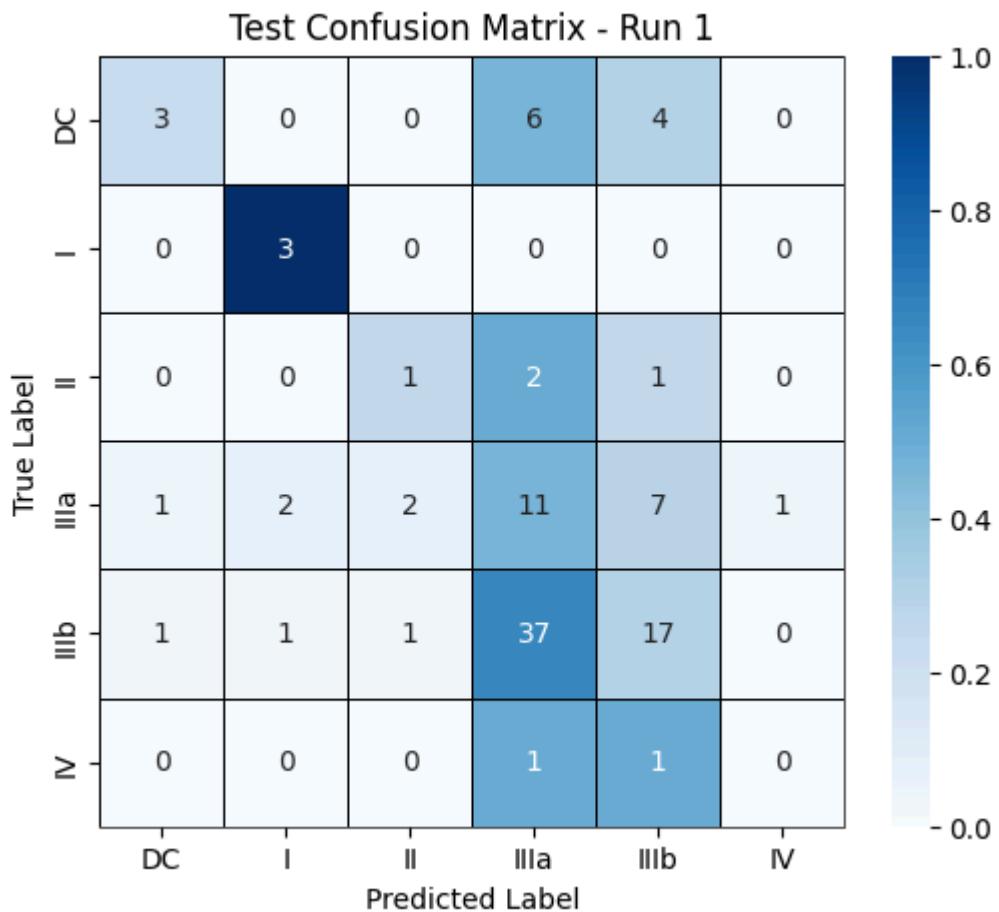
	precision	recall	f1-score	support
DC	0.6000	0.2308	0.3333	13
I	0.5000	1.0000	0.6667	3
II	0.2500	0.2500	0.2500	4
IIIa	0.1930	0.4583	0.2716	24
IIIb	0.5667	0.2982	0.3908	57
IV	0.0000	0.0000	0.0000	2
accuracy			0.3398	103
macro avg	0.3516	0.3729	0.3187	103
weighted avg	0.4586	0.3398	0.3508	103

Balanced Accuracy: 0.37289136302294196
Confusion Matrix:
```
[[ 3  0  0  6  4  0]
 [ 0  3  0  0  0  0]
 [ 0  0  1  2  1  0]
 [ 1  2  2 11  7  1]
 [ 1  1  1 37 17  0]
 [ 0  0  0  1  1  0]]
```

Test Confusion Matrix - Run 1

==
RUN 2: Using Split B for validation, Split A for testing (SWAP)
==
These features will be dropped:
['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ──────────────────────────────────── **0s** 34ms/step

== Test Set Evaluation - Run 2 ==
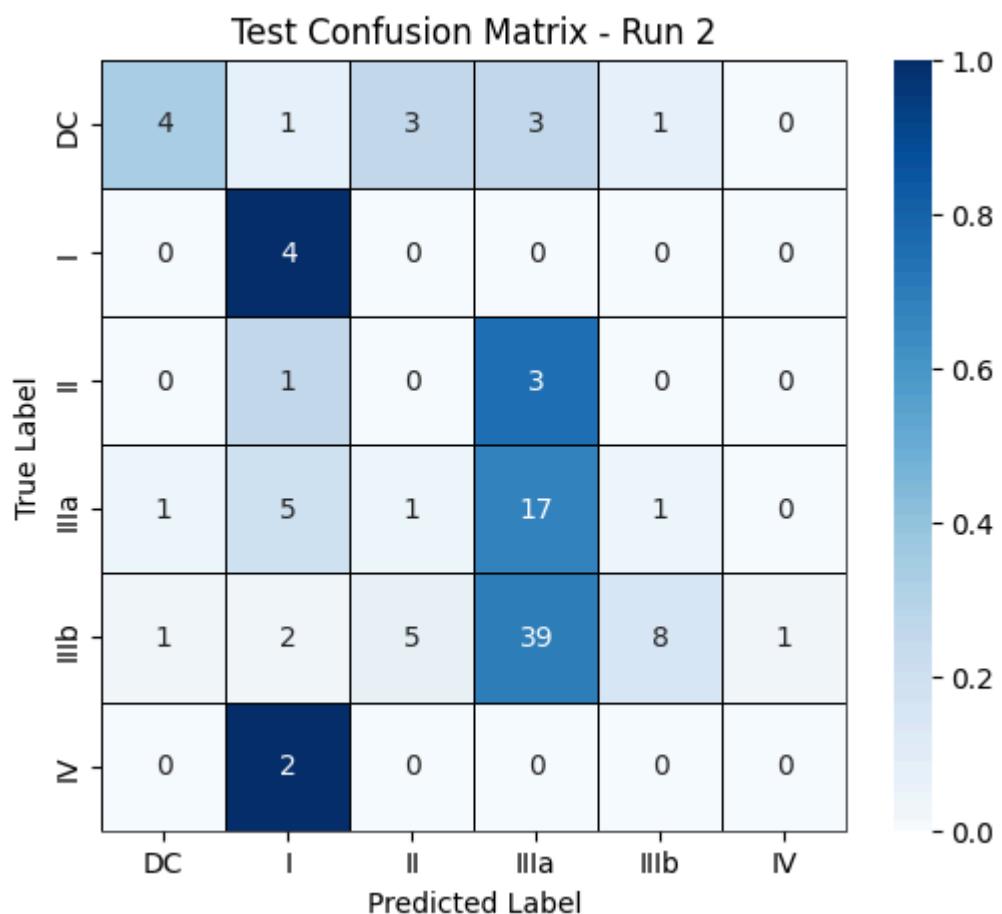
	precision	recall	f1-score	support
DC	0.6667	0.3333	0.4444	12
I	0.2667	1.0000	0.4211	4
II	0.0000	0.0000	0.0000	4
IIIa	0.2742	0.6800	0.3908	25
IIIb	0.8000	0.1429	0.2424	56
IV	0.0000	0.0000	0.0000	2
accuracy			0.3204	103
macro avg	0.3346	0.3594	0.2498	103
weighted avg	0.5895	0.3204	0.2948	103

Balanced Accuracy: 0.35936507936507933

Confusion Matrix:
```
[[ 4  1  3  3  1  0]
 [ 0  4  0  0  0  0]
 [ 0  1  0  3  0  0]
 [ 1  5  1 17  1  0]
 [ 1  2  5 39  8  1]
 [ 0  2  0  0  0  0]]
```

Test Confusion Matrix - Run 2

True Label \ Predicted Label	DC	I	II	IIIa	IIIb	IV
DC	4	1	3	3	1	0
I	0	4	0	0	0	0
II	0	1	0	3	0	0
IIIa	1	5	1	17	1	0
IIIb	1	2	5	39	8	1
IV	0	2	0	0	0	0

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==

== Aggregated Classification Report ==

|	precision	recall	f1-score	support
DC	0.6364	0.2800	0.3889	25
I	0.3333	1.0000	0.5000	7
II	0.0769	0.1250	0.0952	8
IIIa	0.2353	0.5714	0.3333	49
IIIb	0.6250	0.2212	0.3268	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.3301	206
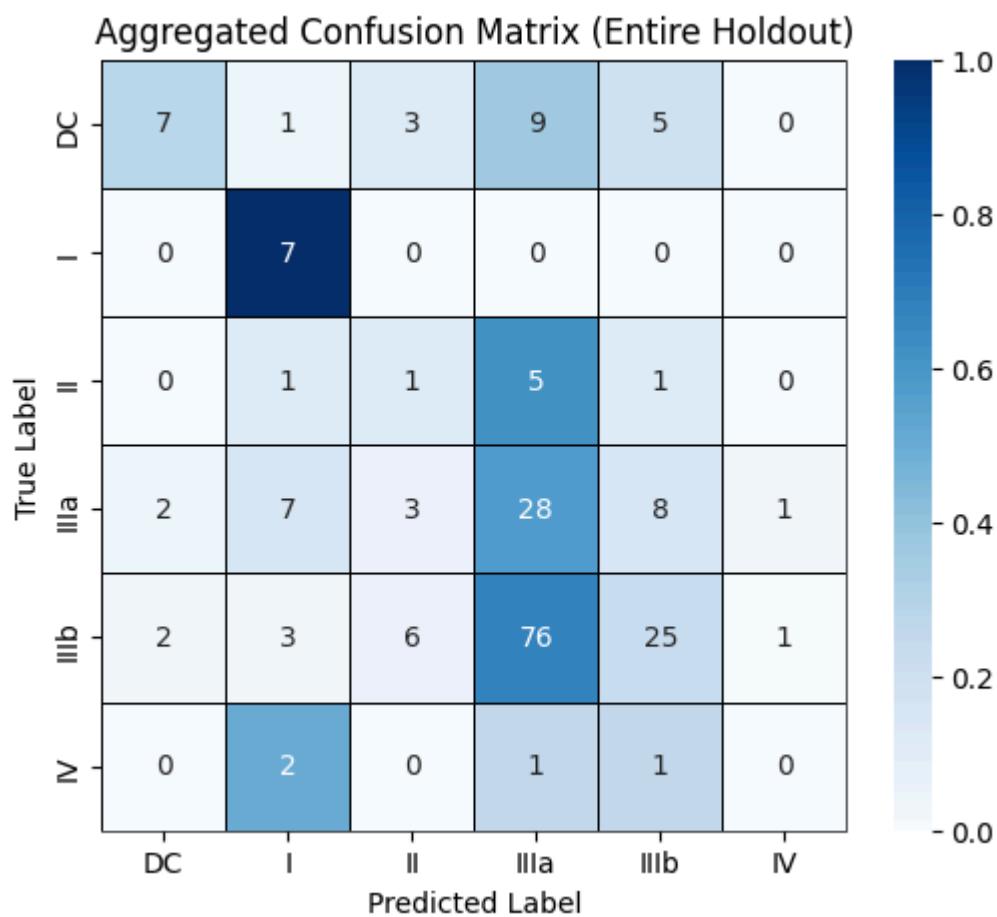
```
  macro avg     0.3178   0.3663   0.2740      206
weighted avg     0.4904   0.3301   0.3264      206
```

Aggregated Balanced Accuracy: 0.3663
Aggregated Confusion Matrix:
```
[[ 7  1  3  9  5  0]
 [ 0  7  0  0  0  0]
 [ 0  1  1  5  1  0]
 [ 2  7  3 28  8  1]
 [ 2  3  6 76 25  1]
 [ 0  2  0  1  1  0]]
```


Aggregated Confusion Matrix (Entire Holdout)

== Average Balanced Accuracy Across Runs ==
Run 1: 0.3729
Run 2: 0.3594
Average: 0.3661

==
RUN 1: Using Split A for validation, Split B for testing
==
These features will be dropped:

['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ━━━━━━━━━━━━━━━━━━━━━━━━━ **0s** 31ms/step

== Test Set Evaluation - Run 1 ==
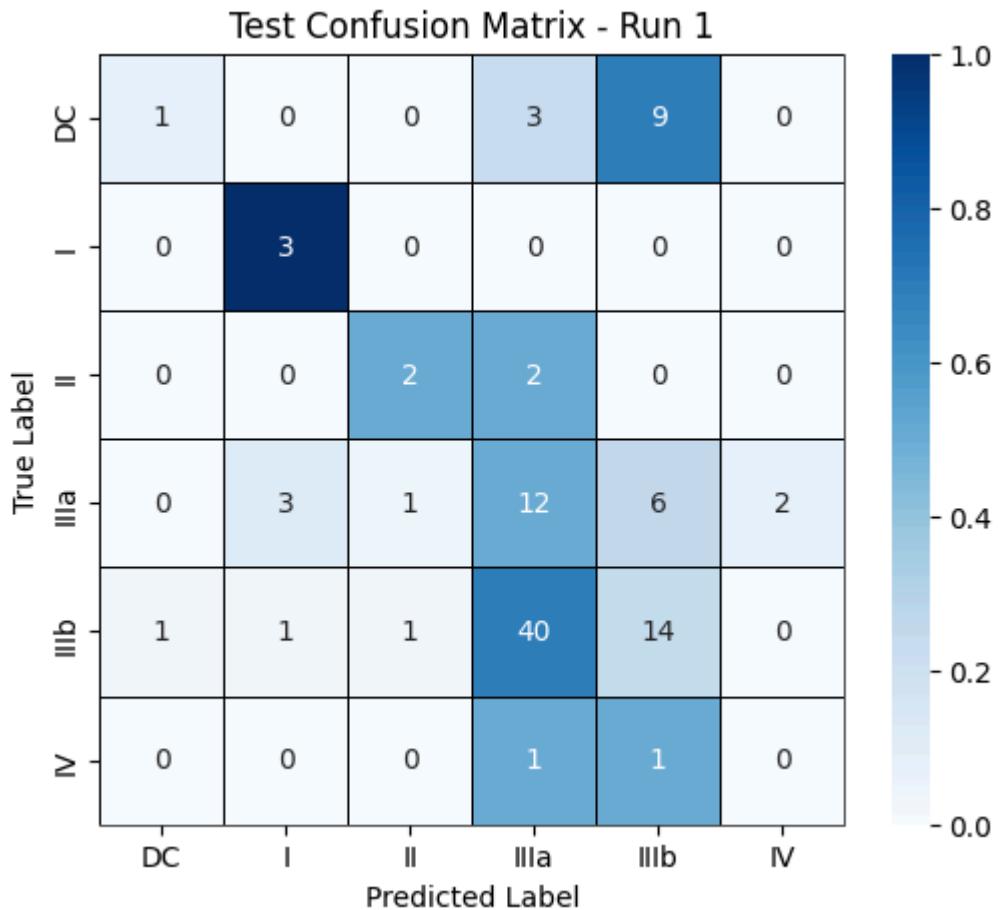
	precision	recall	f1-score	support
DC	0.5000	0.0769	0.1333	13
I	0.4286	1.0000	0.6000	3
II	0.5000	0.5000	0.5000	4
IIIa	0.2069	0.5000	0.2927	24
IIIb	0.4667	0.2456	0.3218	57
IV	0.0000	0.0000	0.0000	2
accuracy			0.3107	103
macro avg	0.3504	0.3871	0.3080	103
weighted avg	0.4015	0.3107	0.3000	103

Balanced Accuracy: 0.38708951866846597
Confusion Matrix:
```
[[ 1  0  0  3  9  0]
 [ 0  3  0  0  0  0]
 [ 0  0  2  2  0  0]
 [ 0  3  1 12  6  2]
 [ 1  1  1 40 14  0]
 [ 0  0  0  1  1  0]]
```

Test Confusion Matrix - Run 1

==
RUN 2: Using Split B for validation, Split A for testing (SWAP)
==
These features will be dropped:
['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ **0s** 20ms/step

== Test Set Evaluation - Run 2 ==
          precision   recall f1-score   support
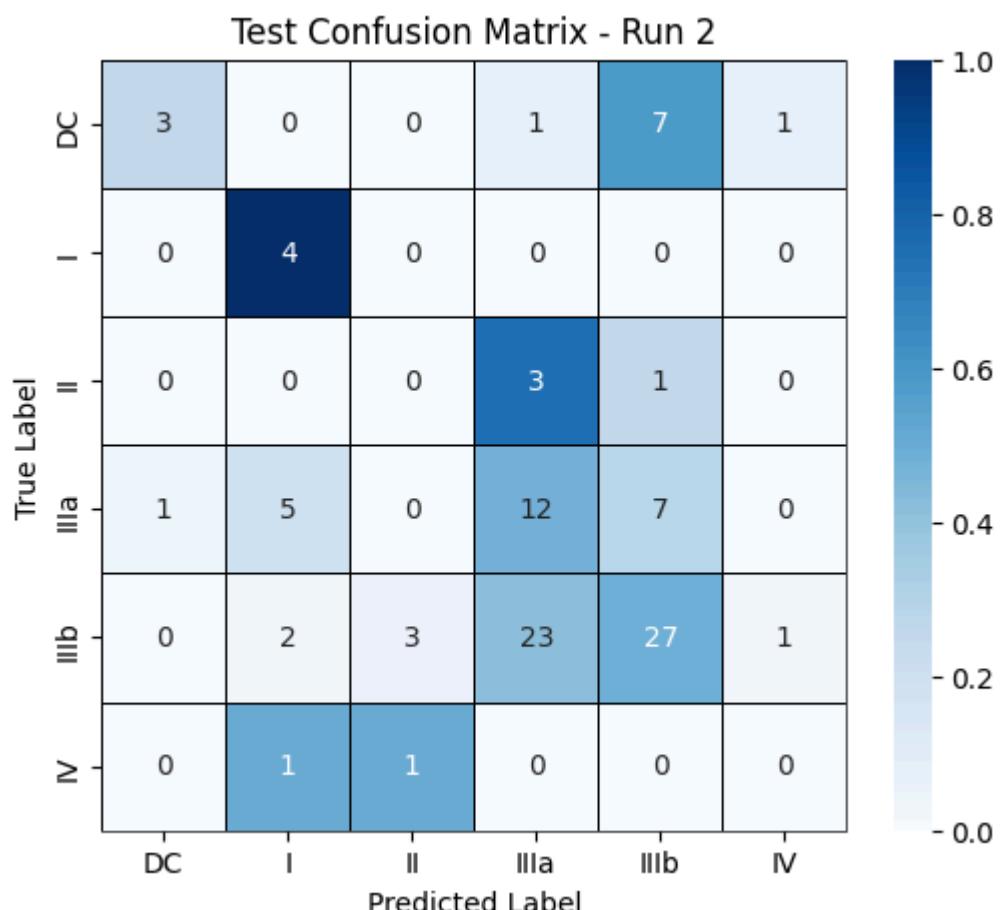
   DC   0.7500   0.2500   0.3750   12
      I   0.3333   1.0000   0.5000   4
     II   0.0000   0.0000   0.0000   4
    IIIa   0.3077   0.4800   0.3750   25
    IIIb   0.6429   0.4821   0.5510   56
     IV   0.0000   0.0000   0.0000   2

  accuracy   0.4466   103
 macro avg   0.3390   0.3687   0.3002   103
weighted avg   0.5245   0.4466   0.4537   103

Balanced Accuracy: 0.3686904761904762

Confusion Matrix:
```
[[ 3  0  0  1  7  1]
 [ 0  4  0  0  0  0]
 [ 0  0  0  3  1  0]
 [ 1  5  0 12  7  0]
 [ 0  2  3 23 27  1]
 [ 0  1  1  0  0  0]]
```

Test Confusion Matrix - Run 2

True Label \ Predicted Label	DC	I	II	IIIa	IIIb	IV
DC	3	0	0	1	7	1
I	0	4	0	0	0	0
II	0	0	0	3	1	0
IIIa	1	5	0	12	7	0
IIIb	0	2	3	23	27	1
IV	0	1	1	0	0	0

==
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
==

== Aggregated Classification Report ==

|	precision	recall	f1-score	support
DC	0.6667	0.1600	0.2581	25
I	0.3684	1.0000	0.5385	7
II	0.2500	0.2500	0.2500	8
IIIa	0.2474	0.4898	0.3288	49
IIIb	0.5694	0.3628	0.4432	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.3786	206
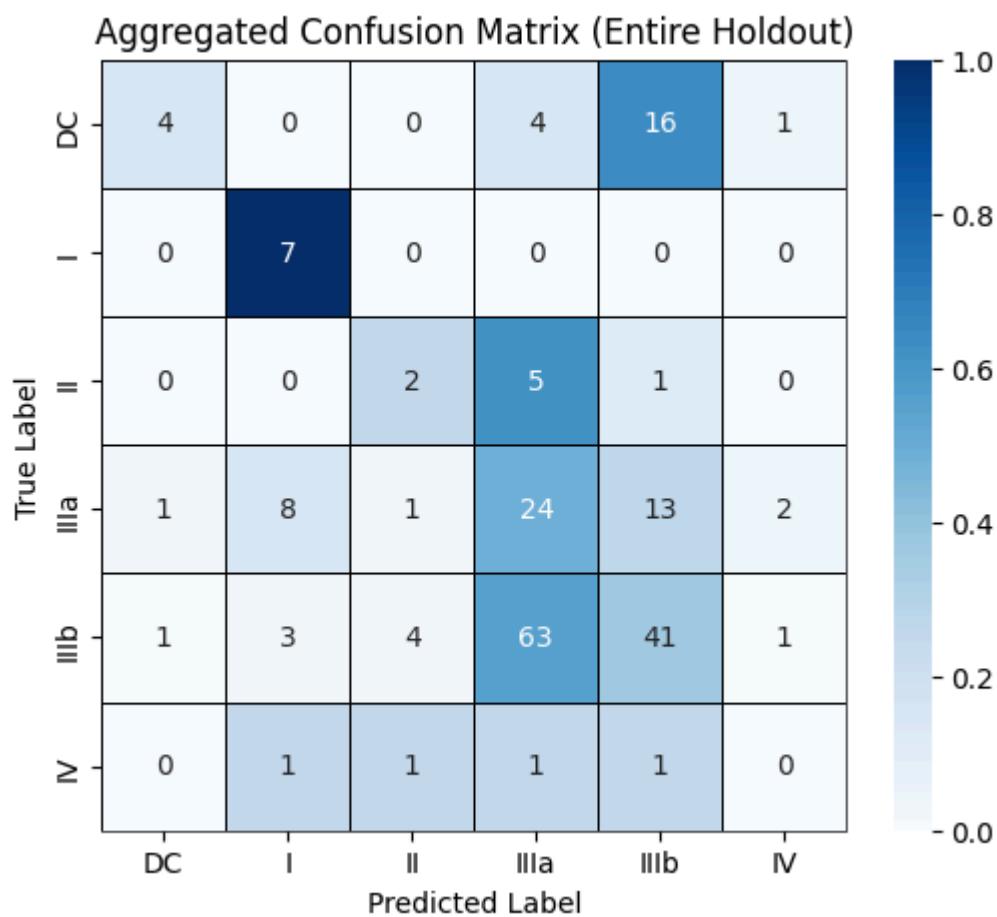
```
    macro avg     0.3503   0.3771   0.3031      206
weighted avg      0.4744   0.3786   0.3807      206
```

Aggregated Balanced Accuracy: 0.3771
Aggregated Confusion Matrix:
```
[[ 4  0  0  4 16  1]
 [ 0  7  0  0  0  0]
 [ 0  0  2  5  1  0]
 [ 1  8  1 24 13  2]
 [ 1  3  4 63 41  1]
 [ 0  1  1  1  1  0]]
```


Aggregated Confusion Matrix (Entire Holdout)

== Average Balanced Accuracy Across Runs ==
Run 1: 0.3871
Run 2: 0.3687
Average: 0.3779

==
RUN 1: Using Split A for validation, Split B for testing
==
These features will be dropped:

['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']
4/4 ━━━━━━━━━━━━━━━━━━━━━━━━━━━ **0s** 29ms/step

== Test Set Evaluation - Run 1 ==

	precision	recall	f1-score	support
DC	0.3500	0.5385	0.4242	13
I	0.3750	1.0000	0.5455	3
II	0.5000	0.2500	0.3333	4
IIIa	0.1714	0.2500	0.2034	24
IIIb	0.6571	0.4035	0.5000	57
IV	0.0000	0.0000	0.0000	2
accuracy			0.3883	103
macro avg	0.3423	0.4070	0.3344	103
weighted avg	0.4781	0.3883	0.4065	103

Balanced Accuracy: 0.4069950517318938
Confusion Matrix:
```
[[ 7  0  0  2  4  0]
 [ 0  3  0  0  0  0]
 [ 1  0  1  1  1  0]
 [ 6  3  0  6  6  3]
 [ 6  2  1 25 23  0]
 [ 0  0  0  1  1  0]]
```

Test Confusion Matrix - Run 1

==
RUN 2: Using Split B for validation, Split A for testing (SWAP)
==
These features will be dropped:
['C113', 'C126', 'C100', 'C66', 'C11', 'C82', 'C135', 'C51', 'C86', 'C85', 'C96', 'C64', 'C87', 'C139', 'C127', 'C101', 'C92', 'C40']

4/4 ──────────────────────────────── **0s** 30ms/step

== Test Set Evaluation - Run 2 ==

	precision	recall	f1-score	support
DC	0.6667	0.1667	0.2667	12
I	0.1739	1.0000	0.2963	4
II	0.0000	0.0000	0.0000	4
IIIa	0.2879	0.7600	0.4176	25
IIIb	0.2000	0.0179	0.0328	56
IV	0.0000	0.0000	0.0000	2
accuracy			0.2524	103
macro avg	0.2214	0.3241	0.1689	103
weighted avg	0.2630	0.2524	0.1618	103

Balanced Accuracy: 0.3240873015873016

Confusion Matrix:
[[2 0 0 4 3 3]
 [0 4 0 0 0 0]
 [0 1 0 2 1 0]
 [0 6 0 19 0 0]
 [1 10 2 41 1 1]
 [0 2 0 0 0 0]]

Test Confusion Matrix - Run 2


```
================================================================
AGGREGATED RESULTS ACROSS ENTIRE HOLDOUT SET
================================================================
```

== Aggregated Classification Report ==

|	precision	recall	f1-score	support
DC	0.3913	0.3600	0.3750	25
I	0.2258	1.0000	0.3684	7
II	0.2500	0.1250	0.1667	8
IIIa	0.2475	0.5102	0.3333	49
IIIb	0.6000	0.2124	0.3137	113
IV	0.0000	0.0000	0.0000	4
accuracy			0.3204	206
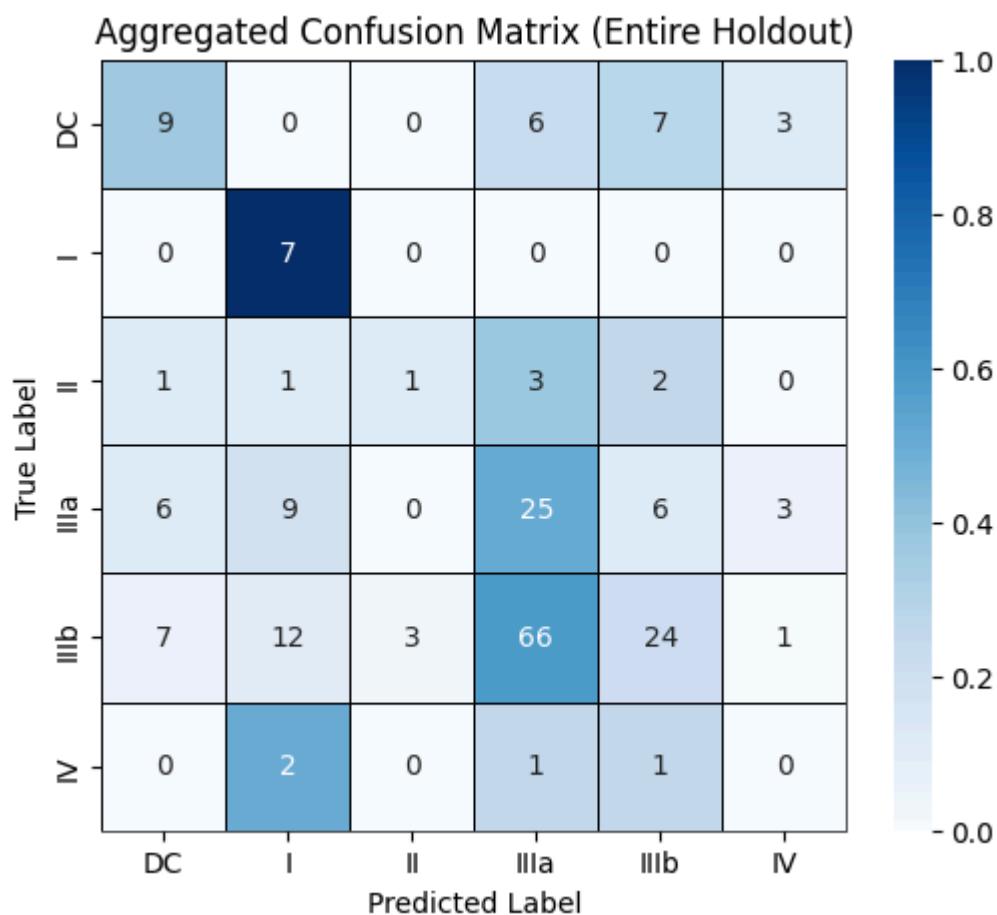
```
   macro avg    0.2858   0.3679   0.2595     206
weighted avg    0.4529   0.3204   0.3159     206
```

Aggregated Balanced Accuracy: 0.3679
Aggregated Confusion Matrix:
```
[[ 9  0  0  6  7  3]
 [ 0  7  0  0  0  0]
 [ 1  1  1  3  2  0]
 [ 6  9  0 25  6  3]
 [ 7 12  3 66 24  1]
 [ 0  2  0  1  1  0]]
```


Aggregated Confusion Matrix (Entire Holdout)

== Average Balanced Accuracy Across Runs ==
Run 1: 0.4070
Run 2: 0.3241
Average: 0.3655

[0.3562, 0.4015, 0.3663, 0.3771, 0.3679]

0.373 ± 0.009