

===== CV RUN 1/5 =====

Starting 5-split CV (3 train / 1 val / 1 test) on 398 samples...

===== Fold 1 =====

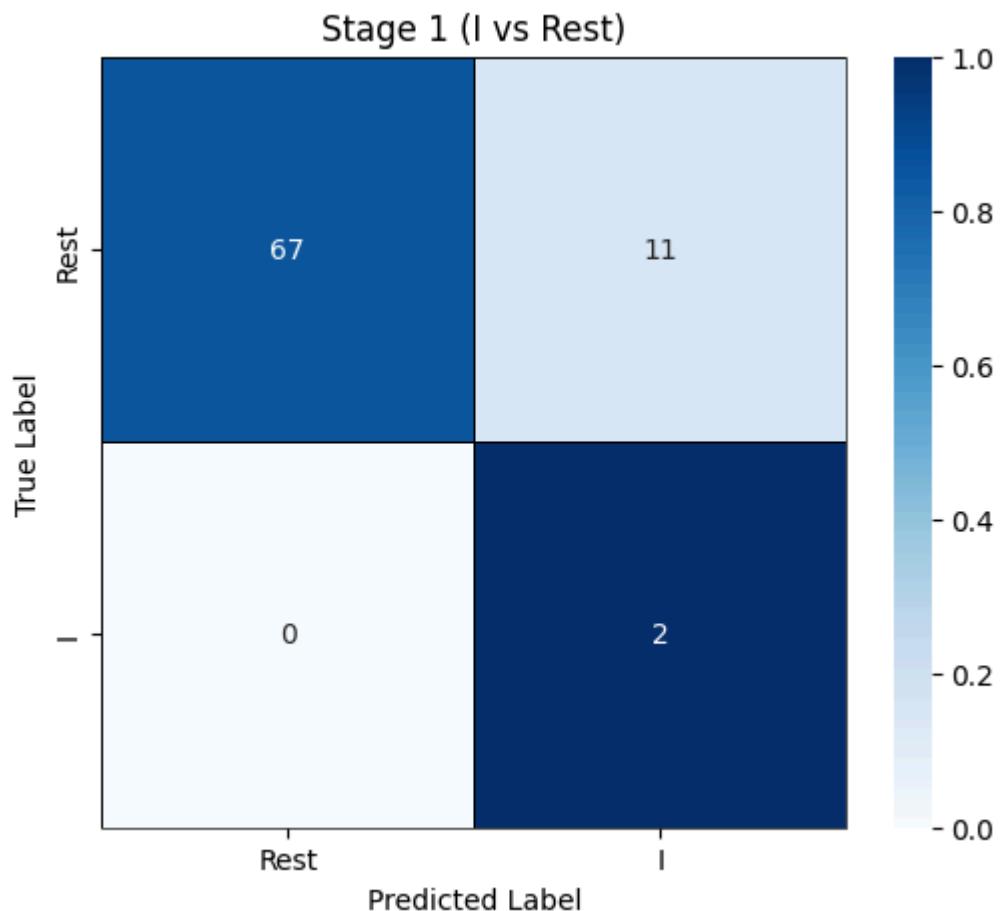
Train: 238, Val: 80, Test: 80

These features will be dropped:

['C11', 'C101', 'C64', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']

-- Stage 1 (I vs Rest) --

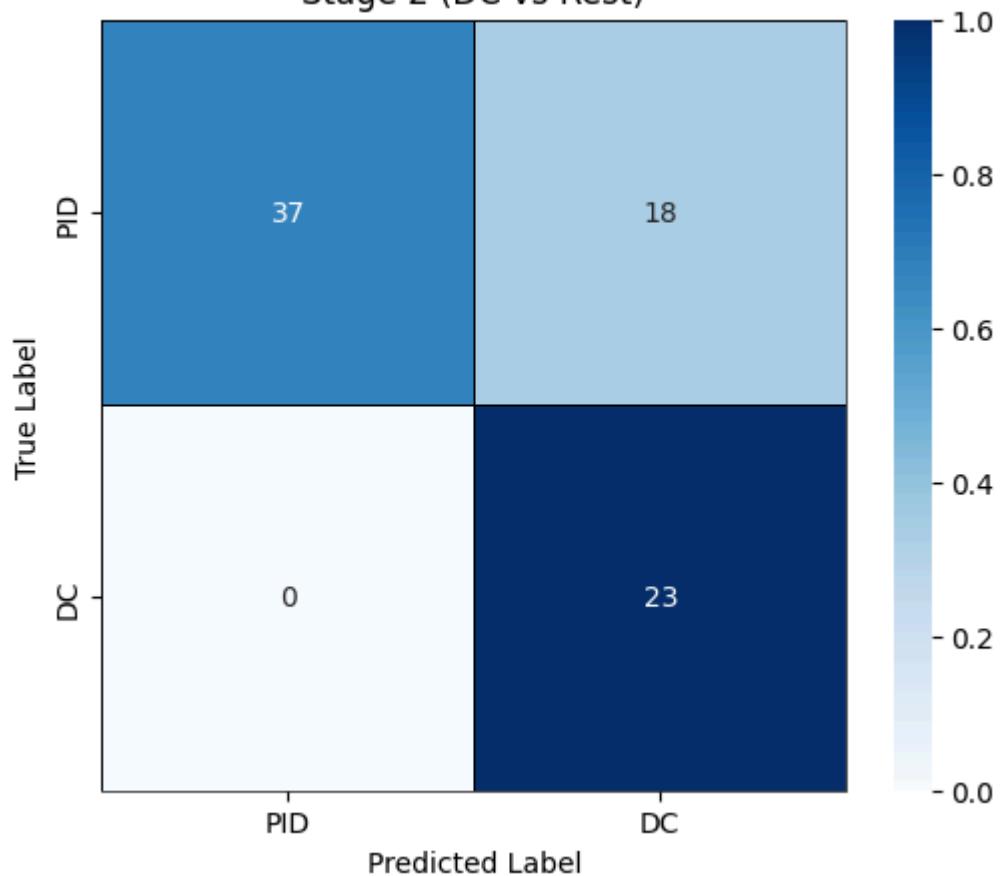
[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.688



-- Stage 2 (DC vs Rest) --

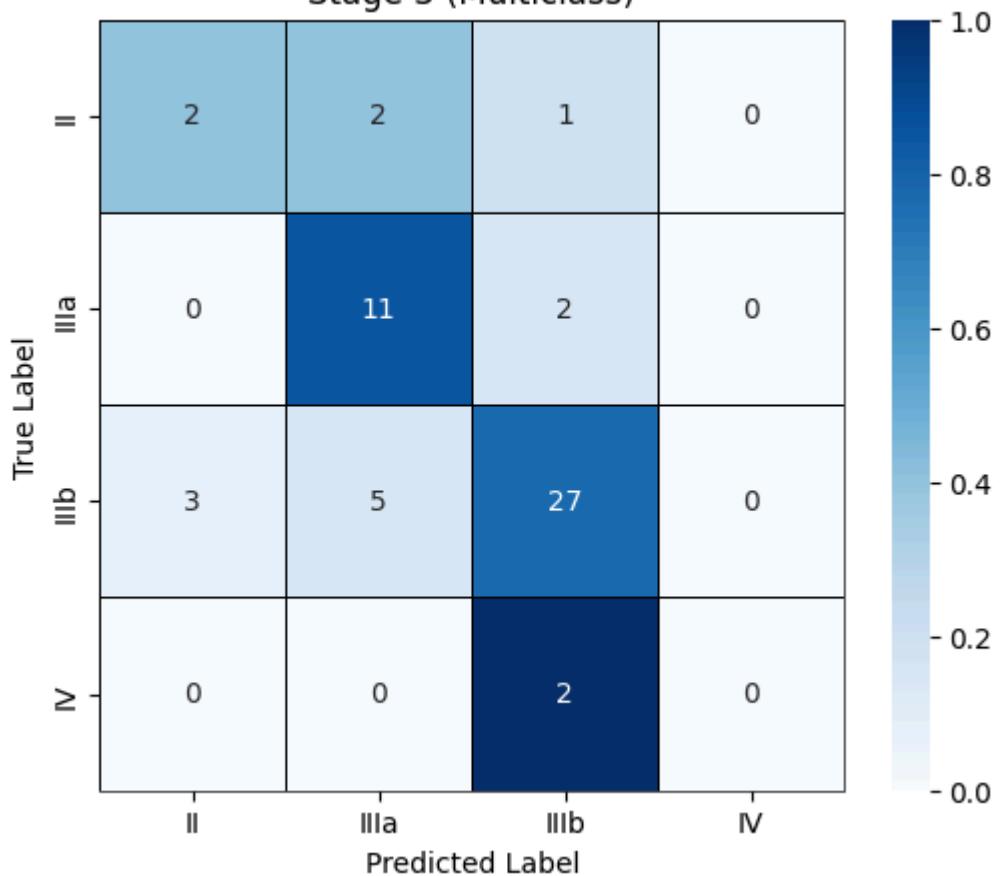
[Threshold Optimization] Best balanced_accuracy: 0.8390 at threshold=0.166

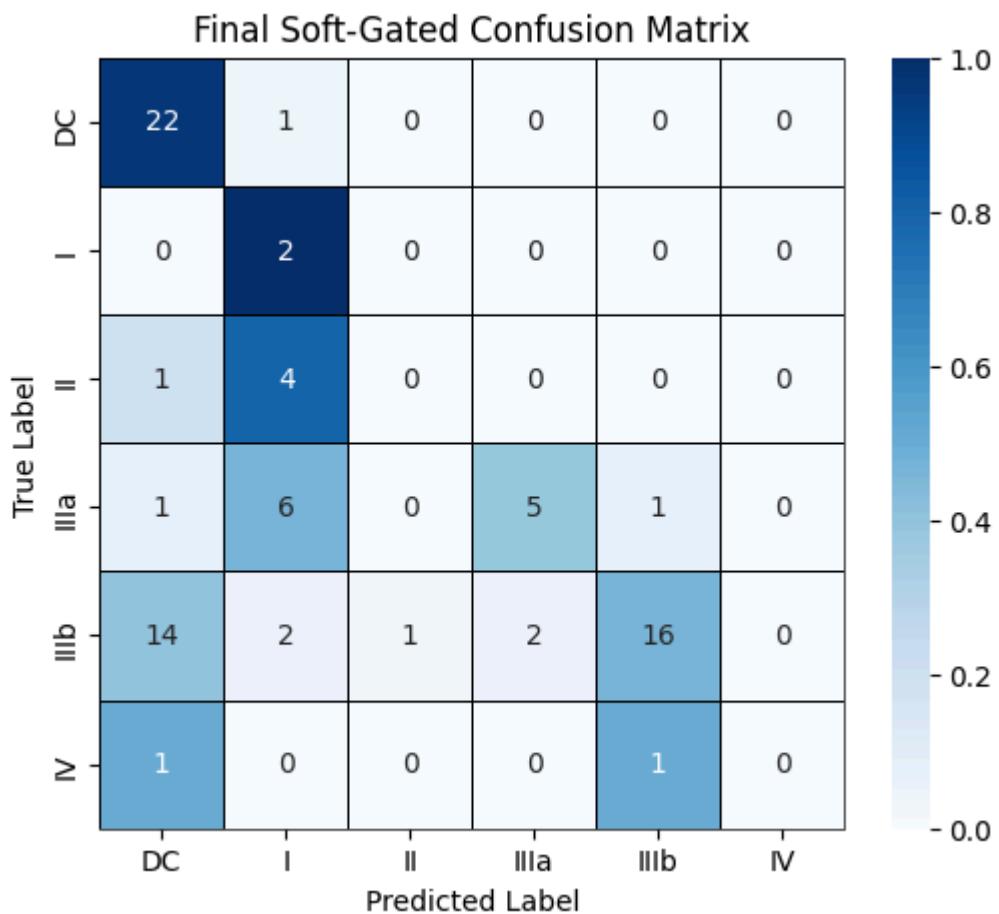
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 2 =====

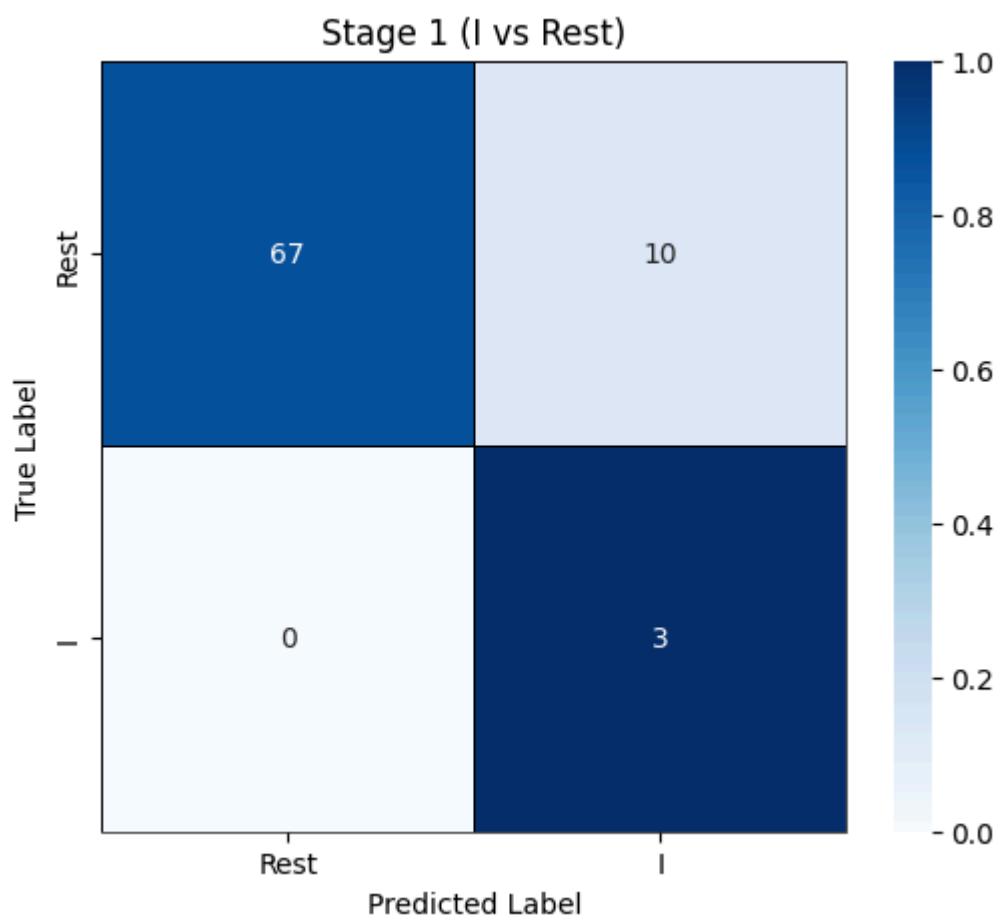
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

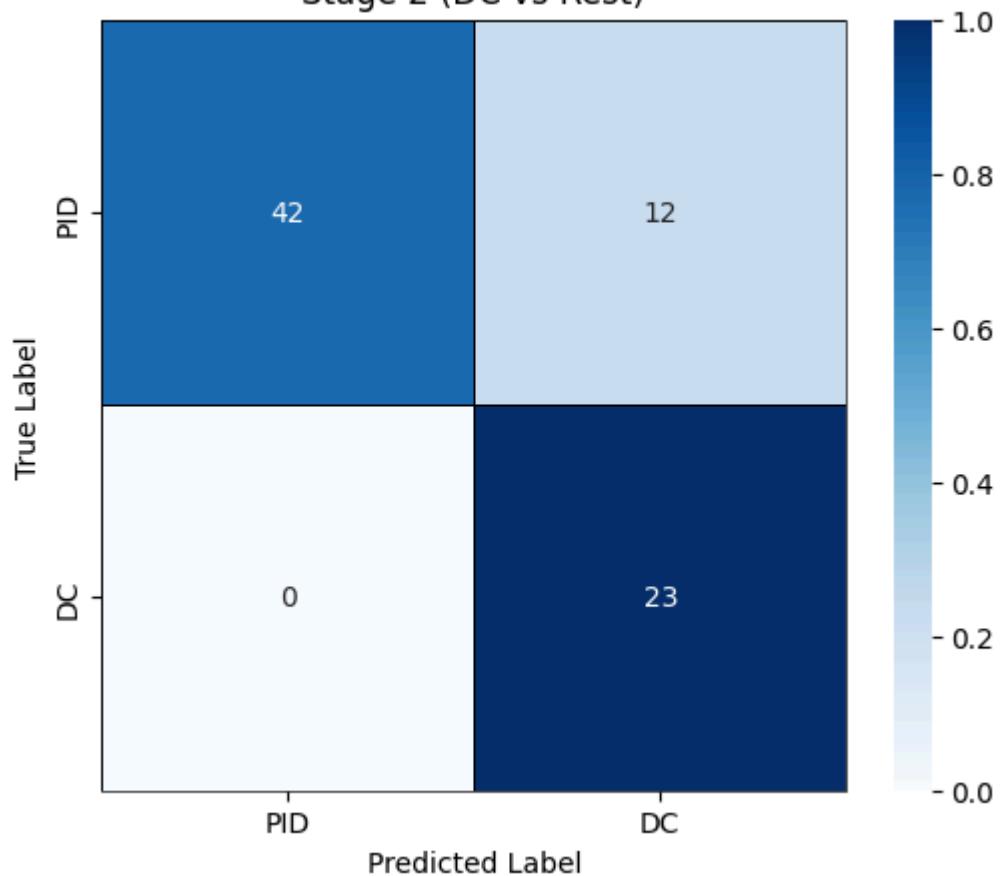
[Threshold Optimization] Best balanced_accuracy: 0.9936 at threshold=0.598



-- Stage 2 (DC vs Rest) --

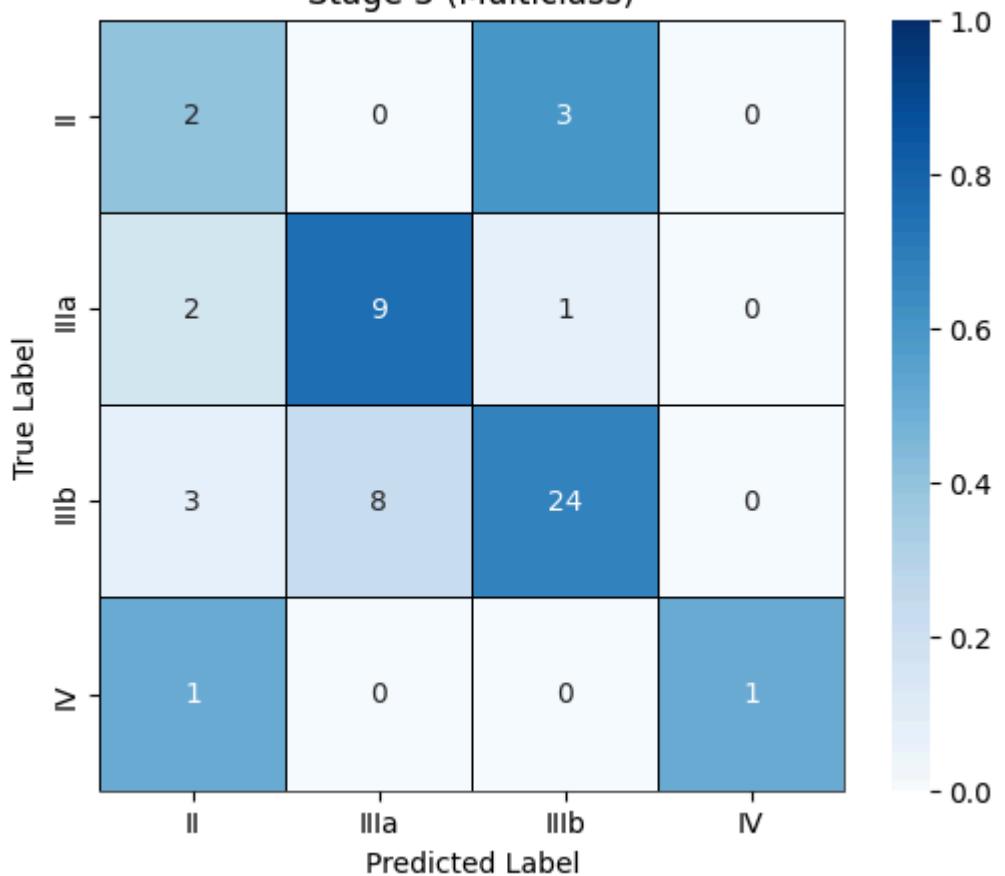
[Threshold Optimization] Best balanced_accuracy: 0.8494 at threshold=0.241

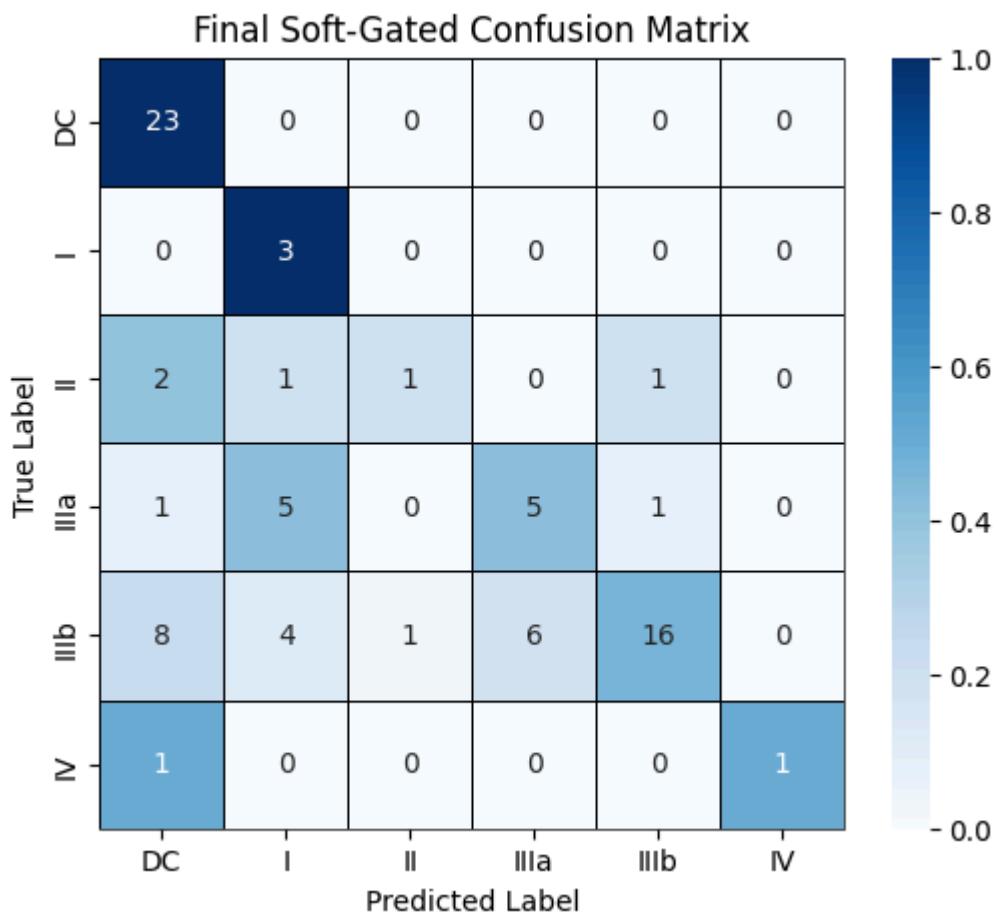
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 3 =====

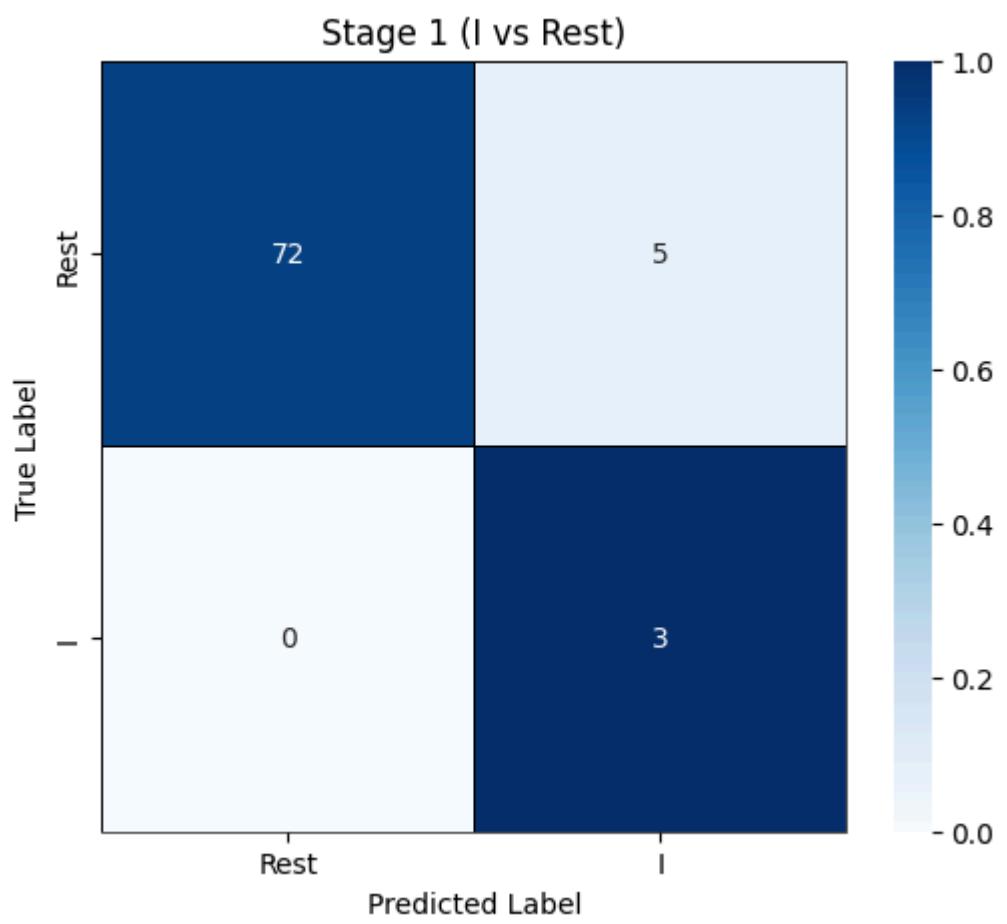
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'MC6. DP T cells', 'C64', 'C100', 'C126', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

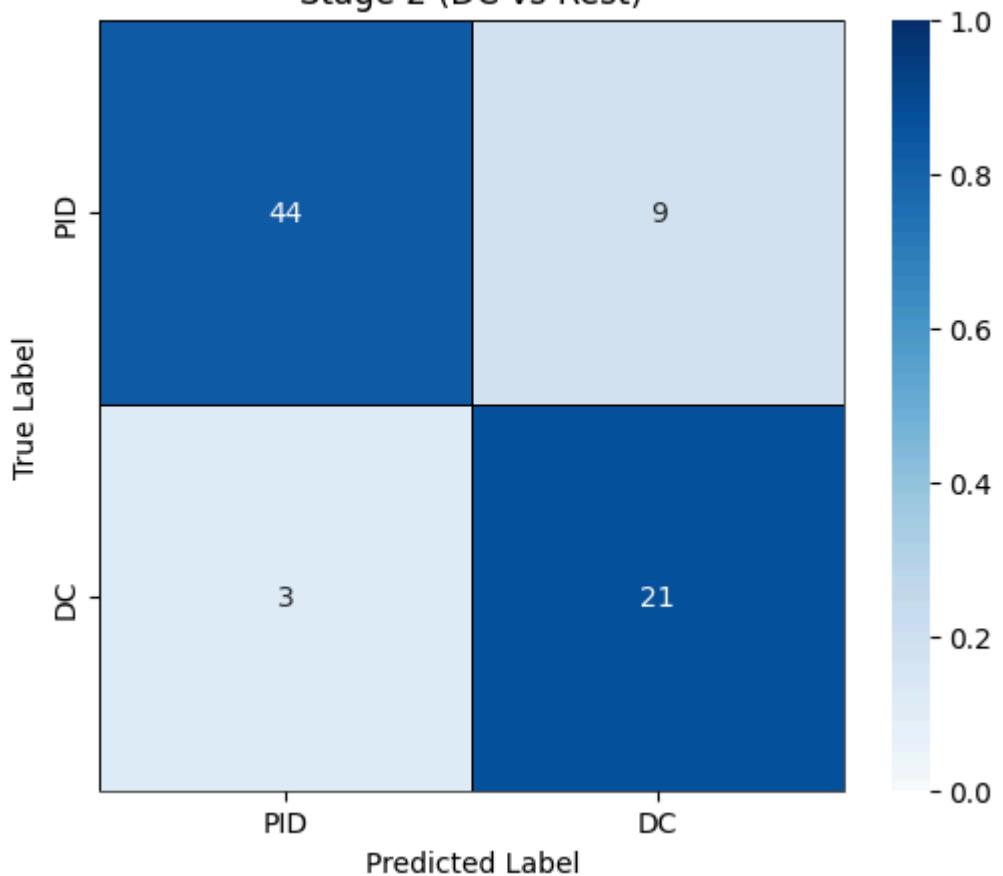
[Threshold Optimization] Best balanced_accuracy: 0.9679 at threshold=0.307



-- Stage 2 (DC vs Rest) --

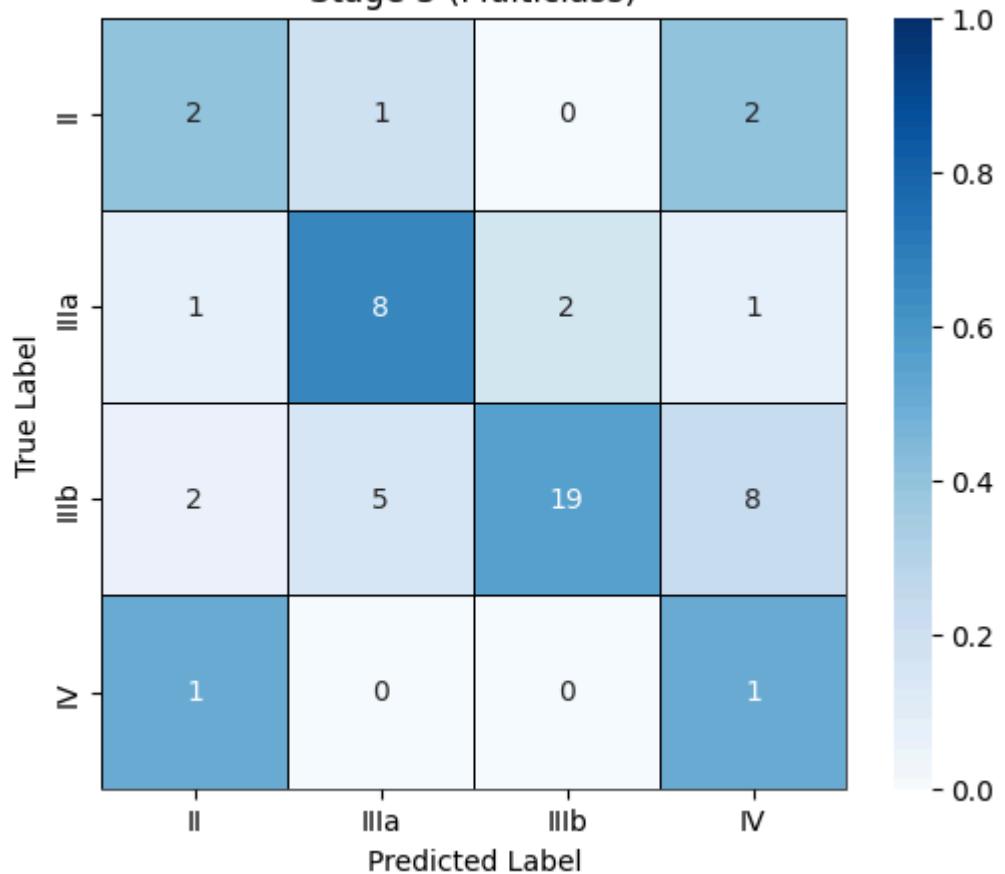
[Threshold Optimization] Best balanced_accuracy: 0.8202 at threshold=0.151

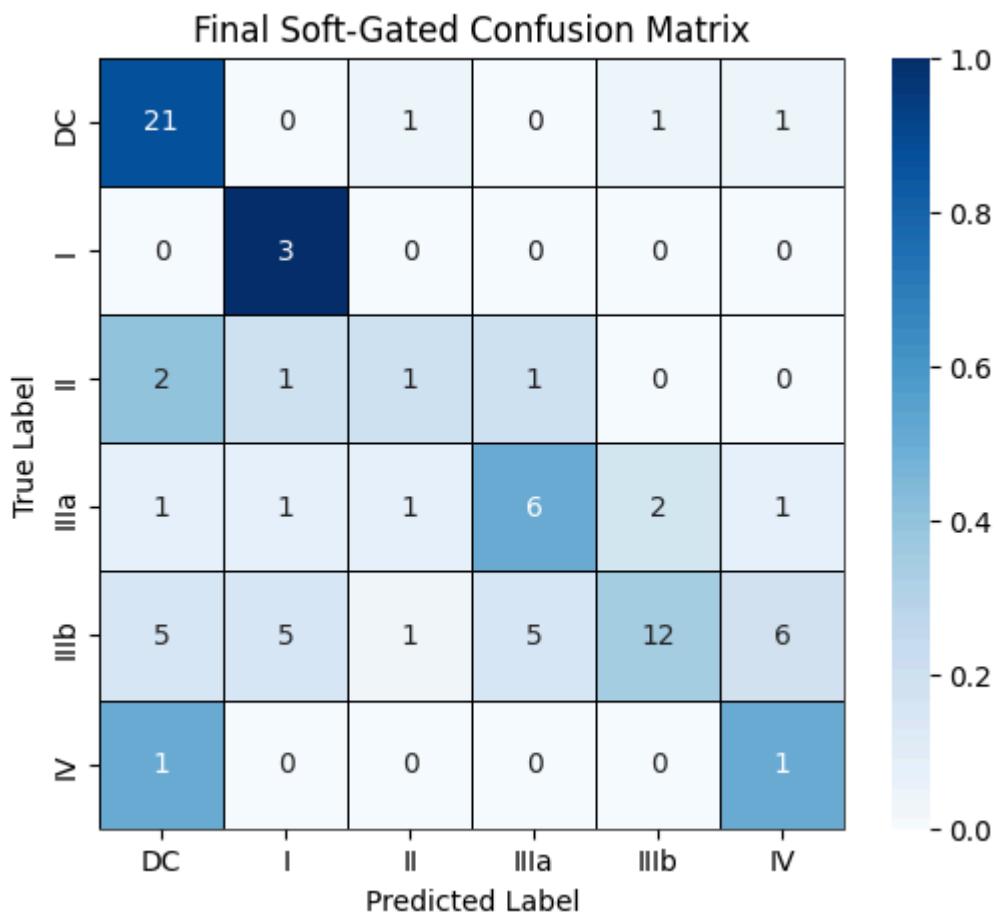
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 4 =====

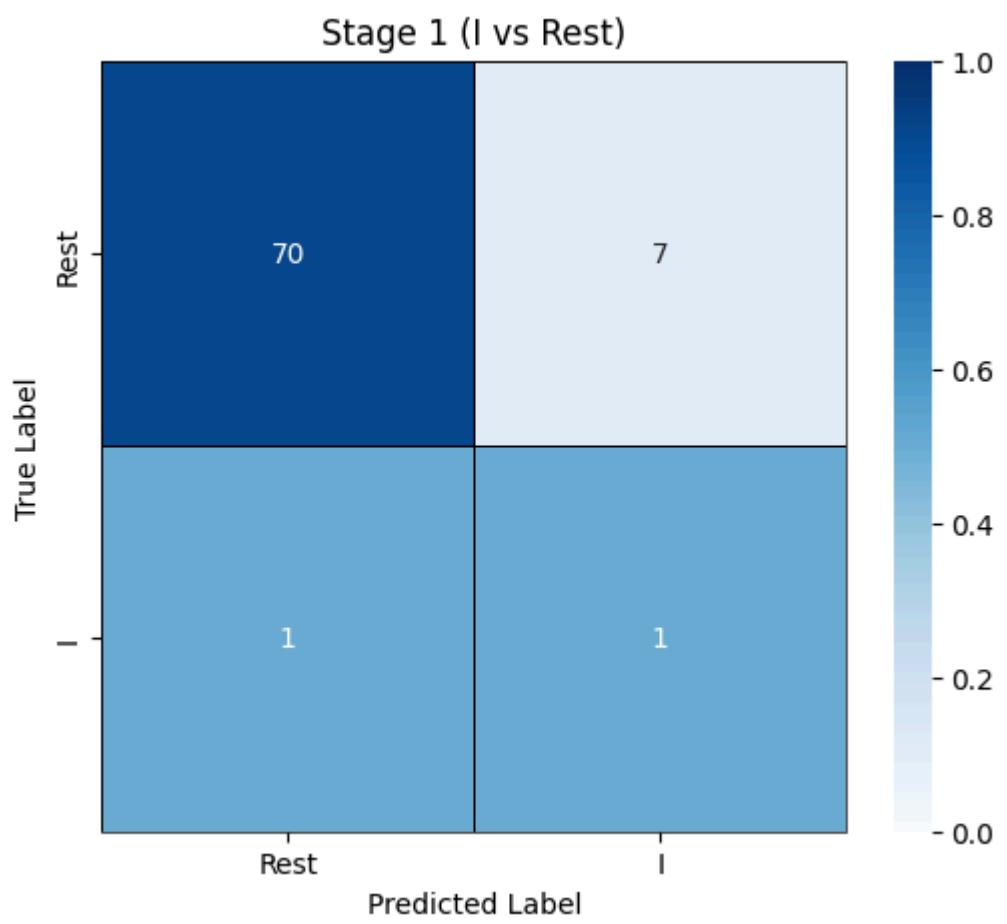
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

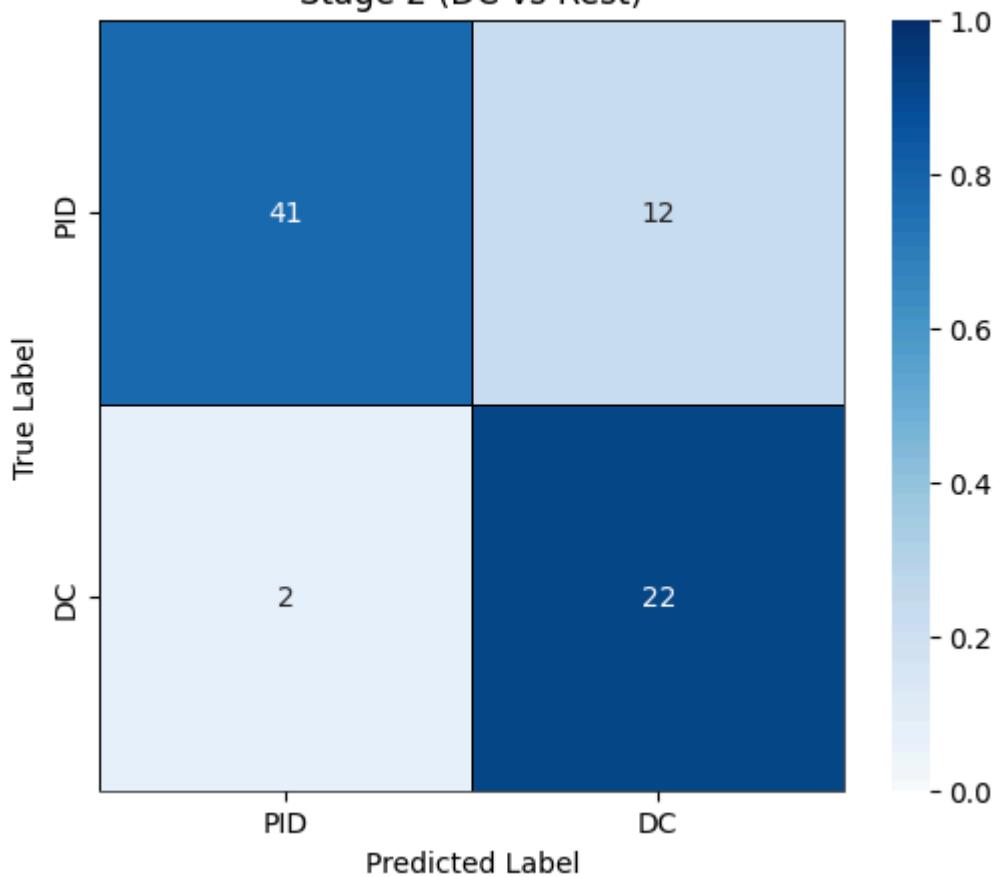
[Threshold Optimization] Best balanced_accuracy: 0.9359 at threshold=0.296



-- Stage 2 (DC vs Rest) --

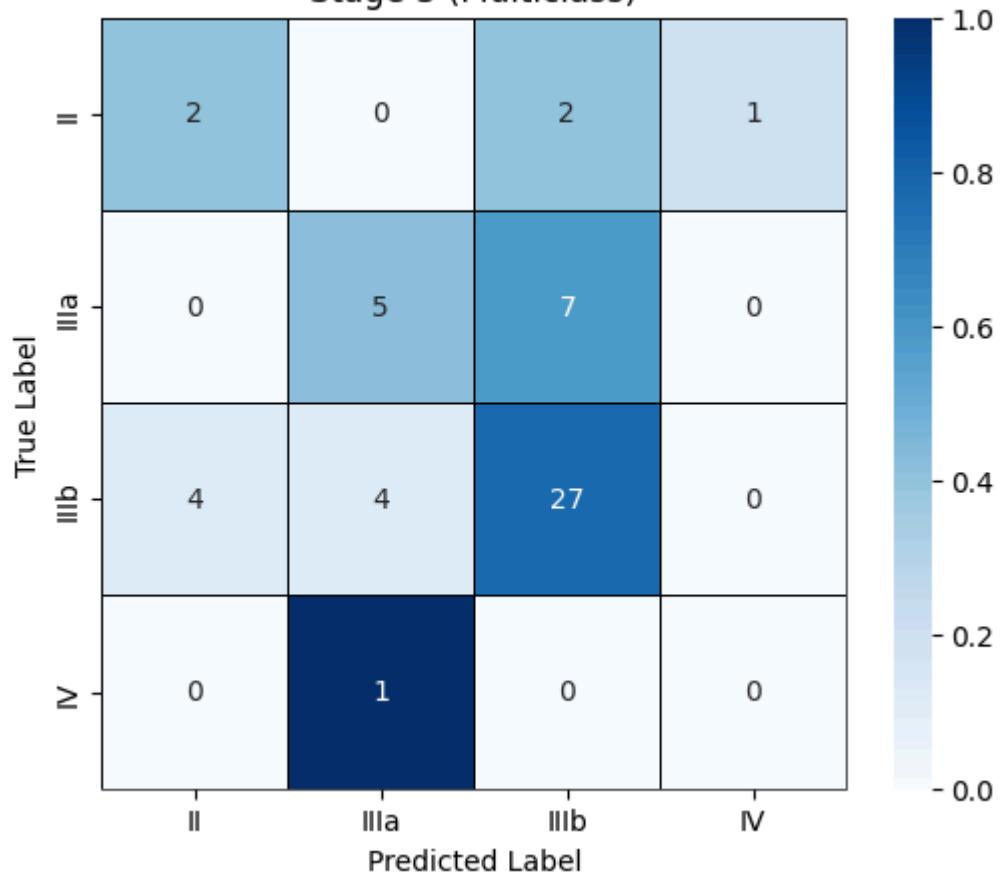
[Threshold Optimization] Best balanced_accuracy: 0.8455 at threshold=0.181

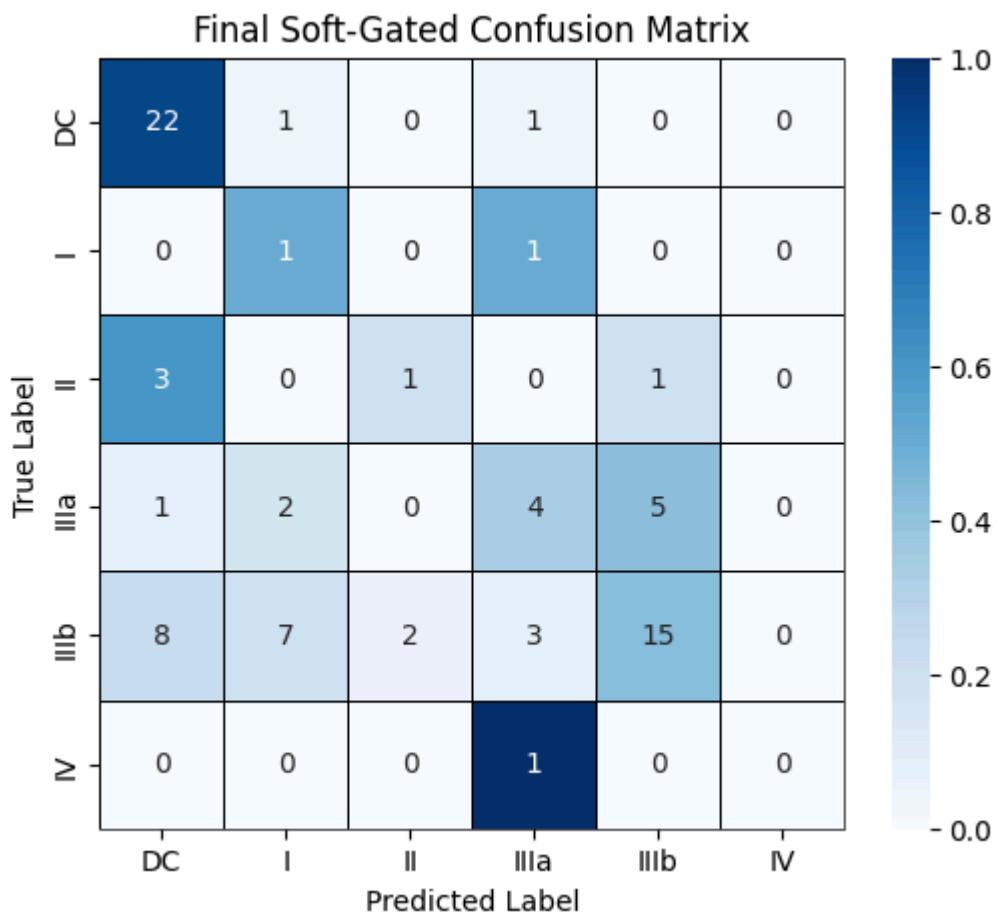
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 5 =====

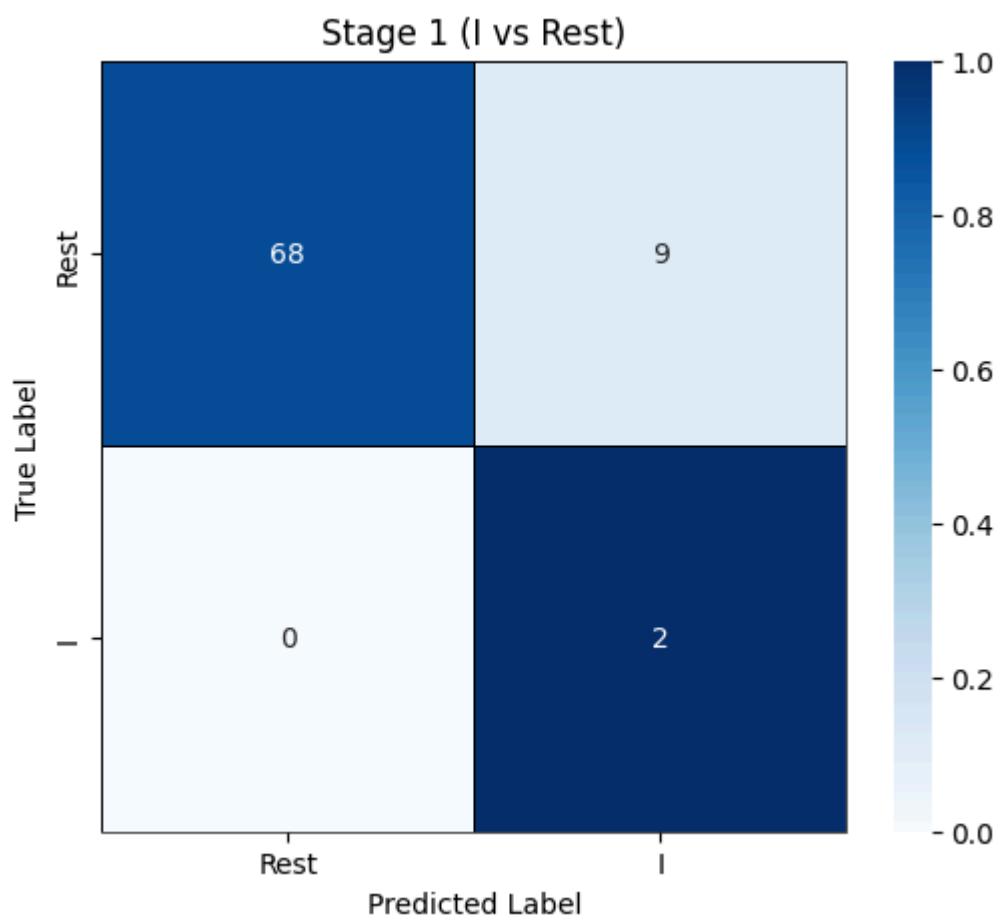
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

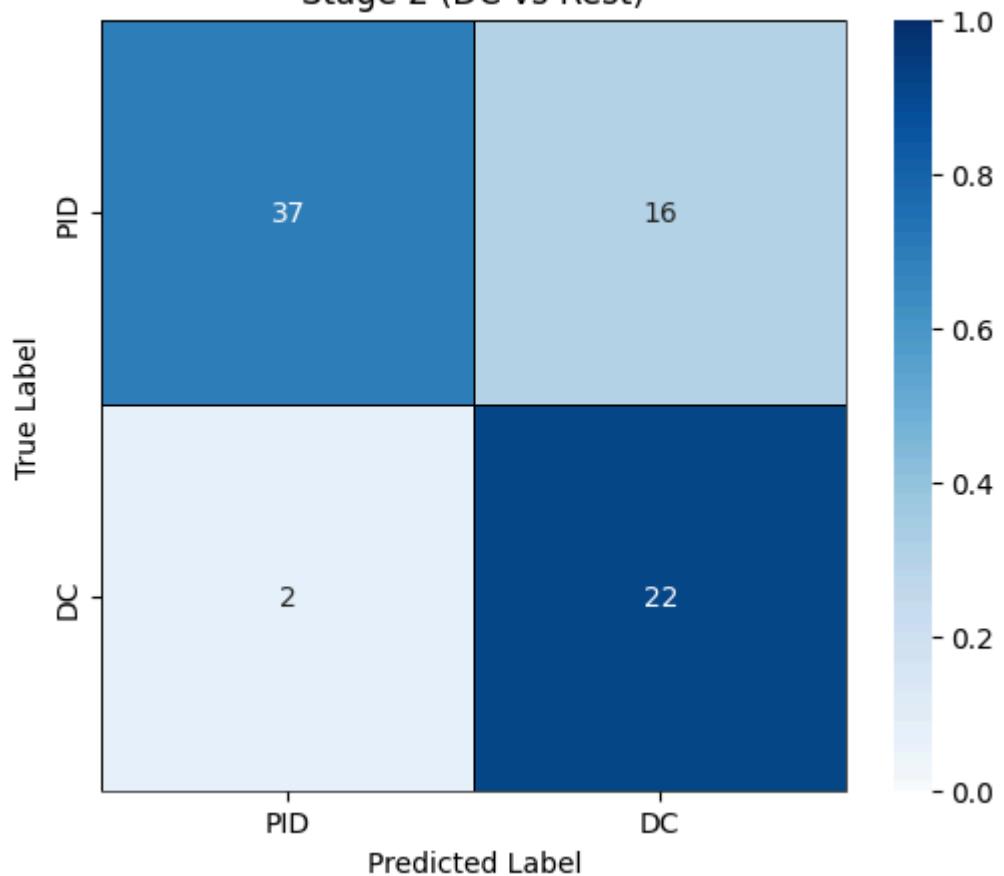
[Threshold Optimization] Best balanced_accuracy: 0.9872 at threshold=0.538



-- Stage 2 (DC vs Rest) --

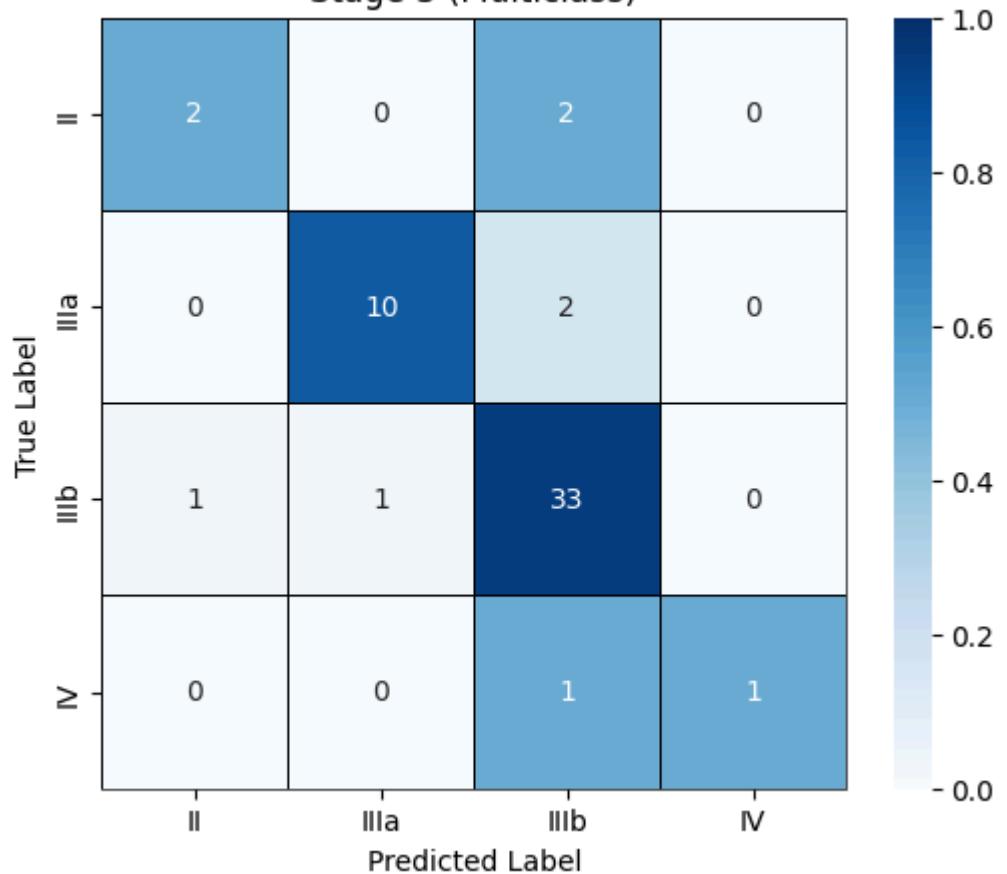
[Threshold Optimization] Best balanced_accuracy: 0.8747 at threshold=0.236

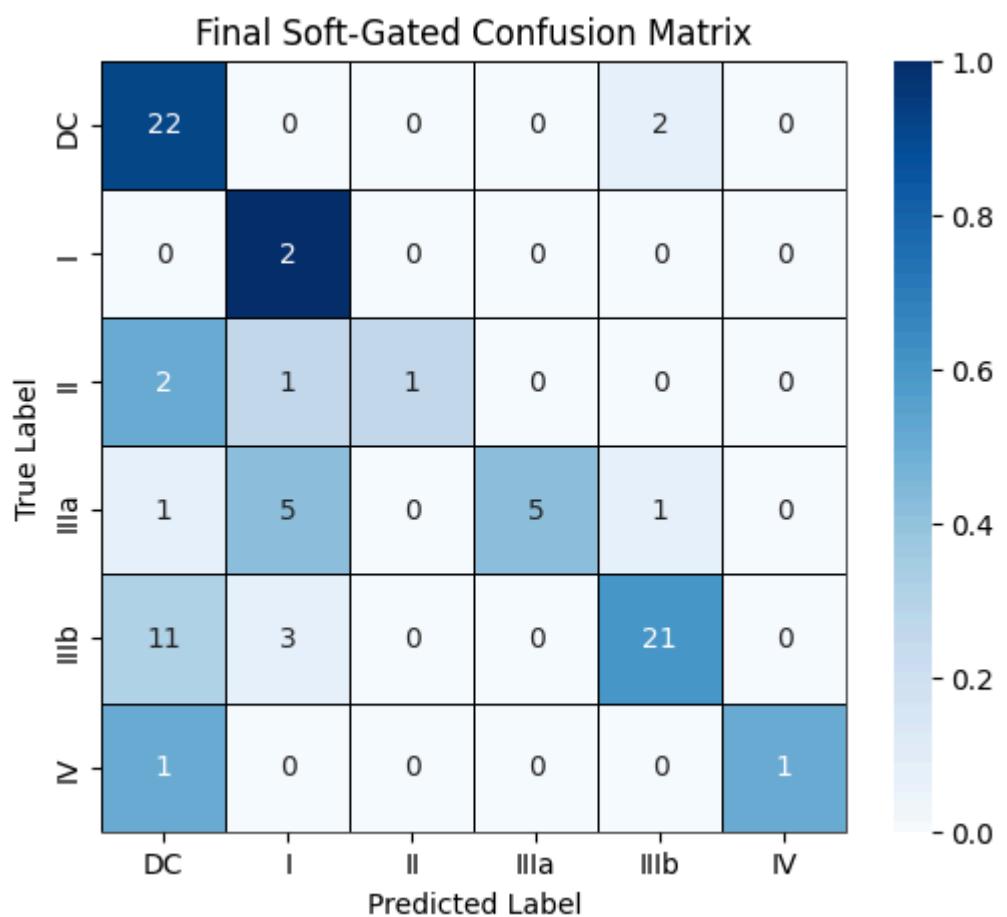
Stage 2 (DC vs Rest)



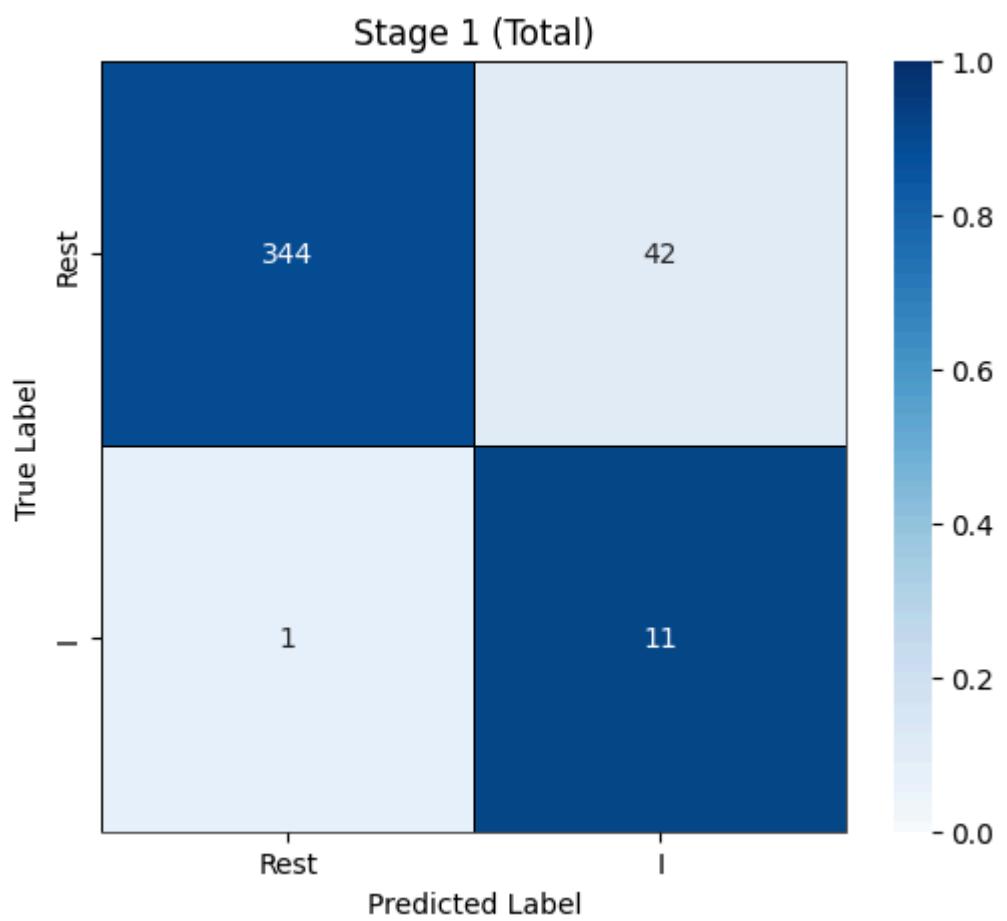
-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)

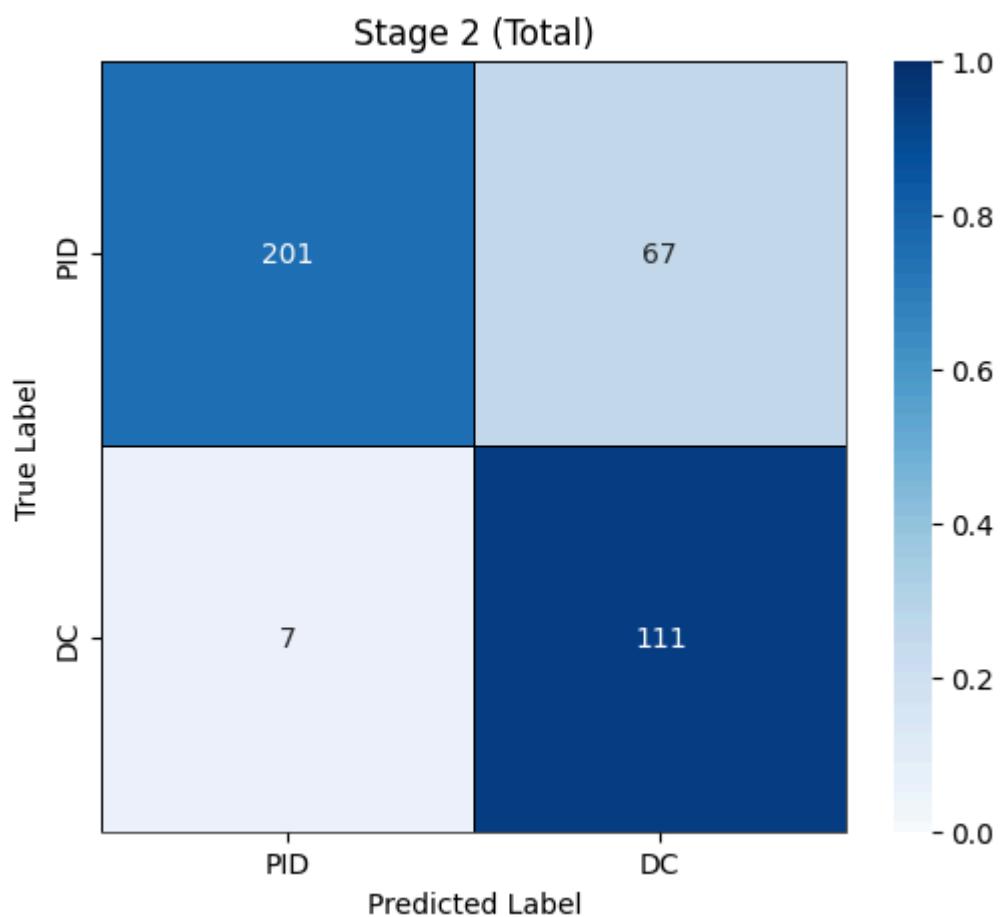




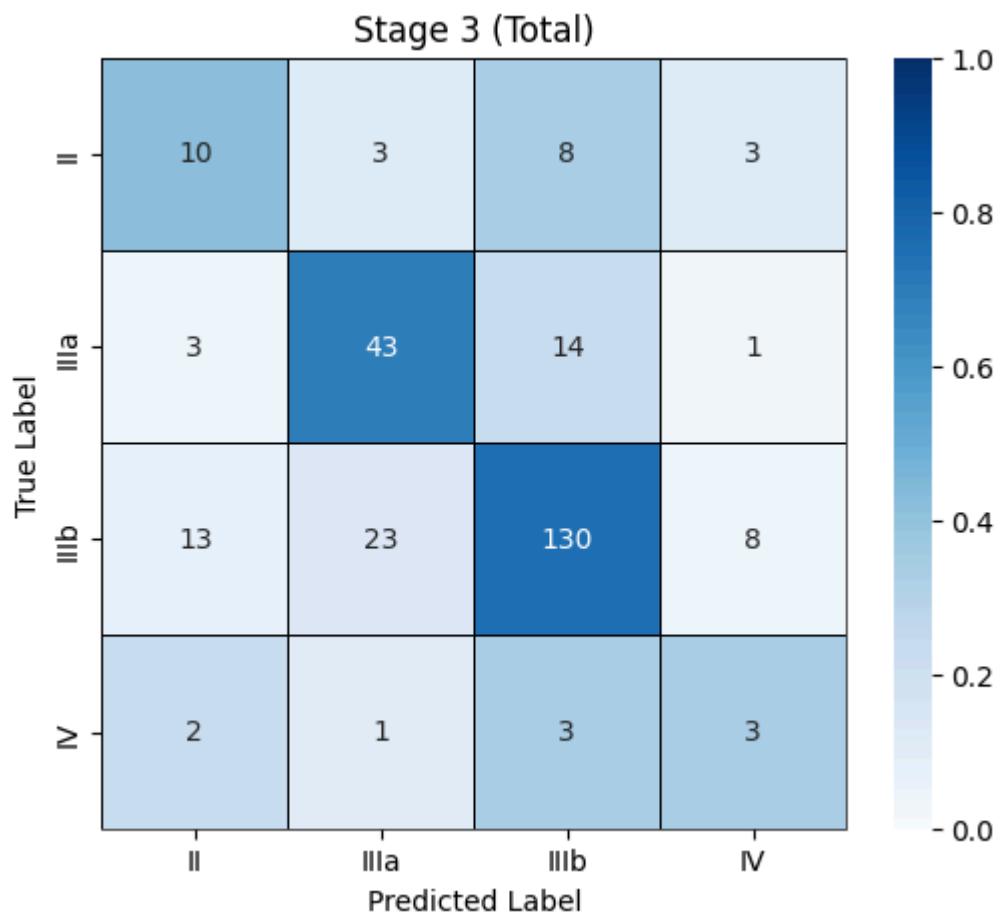
===== AGGREGATED CONFUSION MATRICES =====



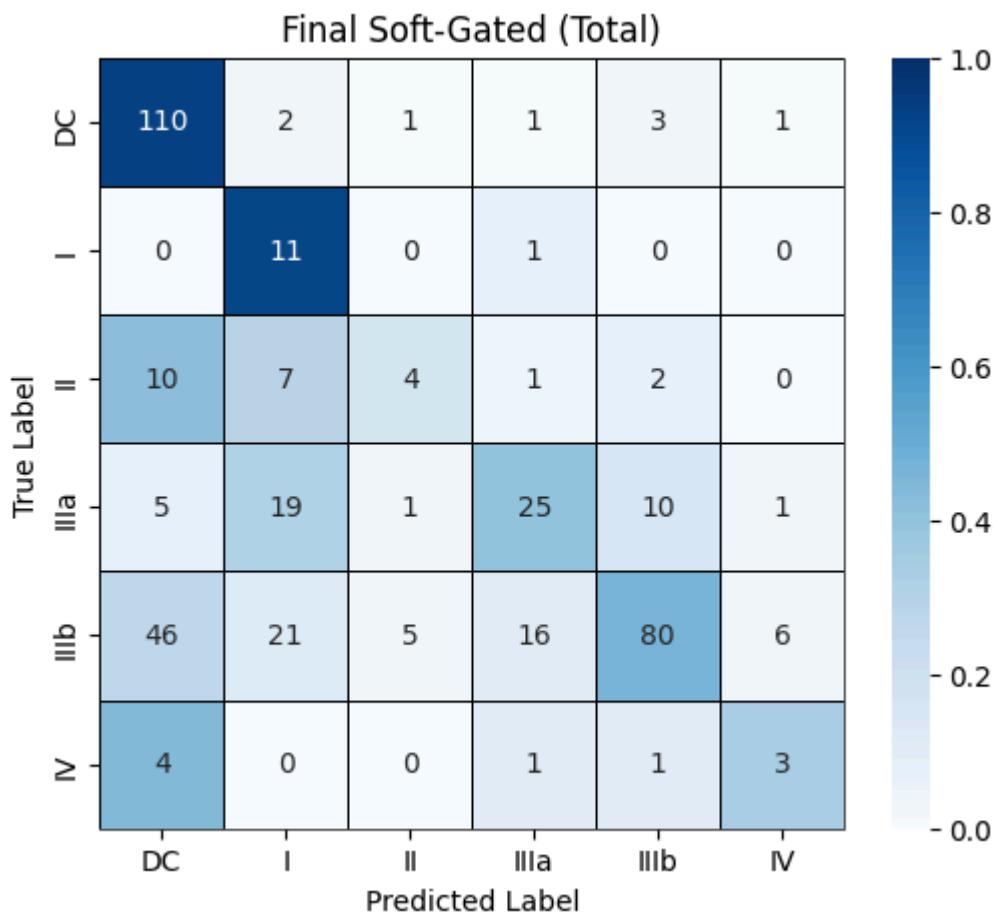
	precision	recall	f1-score	support
0	1.00	0.89	0.94	386
1	0.21	0.92	0.34	12
accuracy		0.89	0.89	398
macro avg	0.60	0.90	0.64	398
weighted avg	0.97	0.89	0.92	398



	precision	recall	f1-score	support
0	0.966	0.750	0.845	268
1	0.624	0.941	0.750	118
accuracy		0.808	0.808	386
macro avg	0.795	0.845	0.797	386
weighted avg	0.862	0.808	0.816	386



	precision	recall	f1-score	support
0	0.357	0.417	0.385	24
1	0.614	0.705	0.656	61
2	0.839	0.747	0.790	174
3	0.200	0.333	0.250	9
accuracy		0.694	0.694	268
macro avg	0.503	0.551	0.520	268
weighted avg	0.723	0.694	0.705	268



===== CV SUMMARY =====

precision recall f1-score support

DC	0.6286	0.9322	0.7509	118
I	0.1833	0.9167	0.3056	12
II	0.3636	0.1667	0.2286	24
IIIa	0.5556	0.4098	0.4717	61
IIIb	0.8333	0.4598	0.5926	174
IV	0.2727	0.3333	0.3000	9

accuracy		0.5854	398	
macro avg	0.4729	0.5364	0.4415	398
weighted avg	0.6695	0.5854	0.5838	398

Balanced Accuracy (Final Soft-Gated, aggregated): 0.5364127061689125

===== Accuracies =====

Per-fold balanced accuracy:

Stage 1: [0.9294871794871795, 0.935064935064935, 0.9675324675324675, 0.70454545454546, 0.9415584415584415]

Stage 2: [0.8363636363636364, 0.8888888888888888, 0.8525943396226415, 0.845125786163522, 0.8073899371069182]

Stage 3: [0.5043956043956044, 0.5839285714285714, 0.5313725490196078, 0.39702380952380956, 0.694047619047619]

Mean per-fold balanced accuracy: [np.float64(0.8956376956376957), np.float64(0.8460725176291213), np.float64(0.5421536306830423)]

Aggregated balanced accuracy:

Stage 1: 0.9039291882556131

Stage 2: 0.8453389830508475

Stage 3: 0.5505111173921237

Final Soft-Gated: 0.5364127061689125

===== CV RUN 2/5 =====

Starting 5-split CV (3 train / 1 val / 1 test) on 398 samples...

===== Fold 1 =====

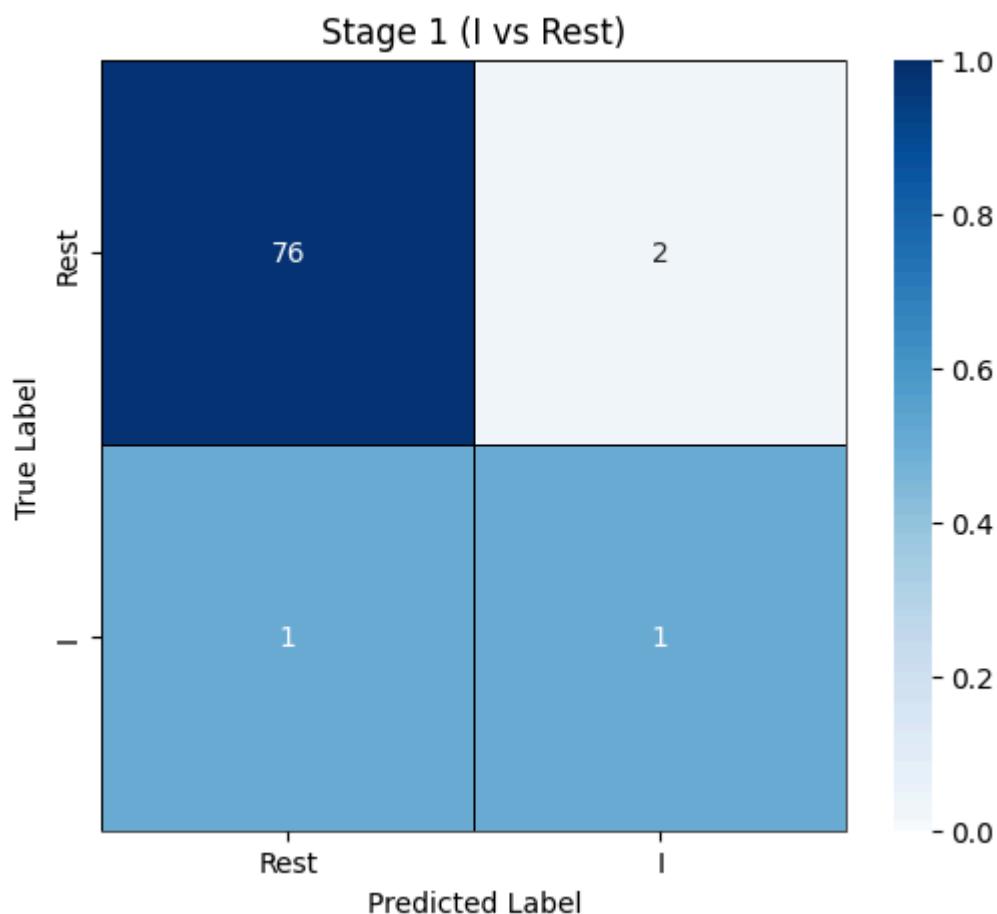
Train: 238, Val: 80, Test: 80

These features will be dropped:

['C11', 'C101', 'C64', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']

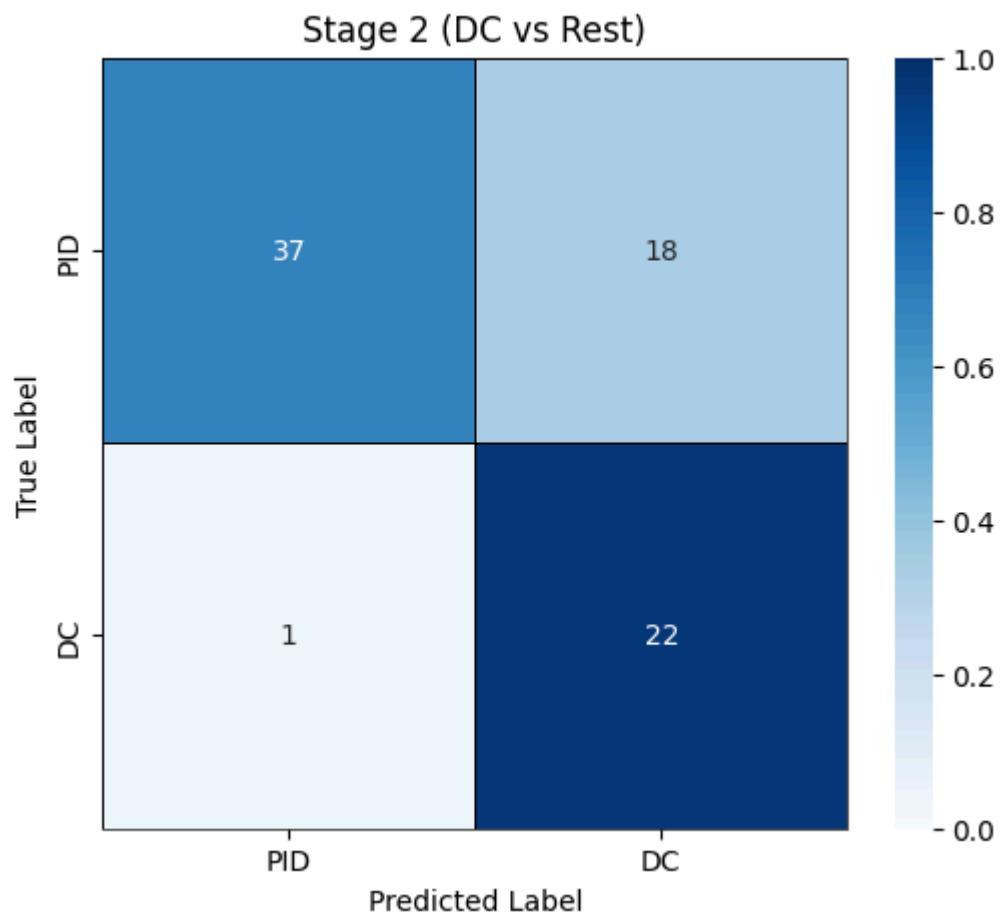
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.352



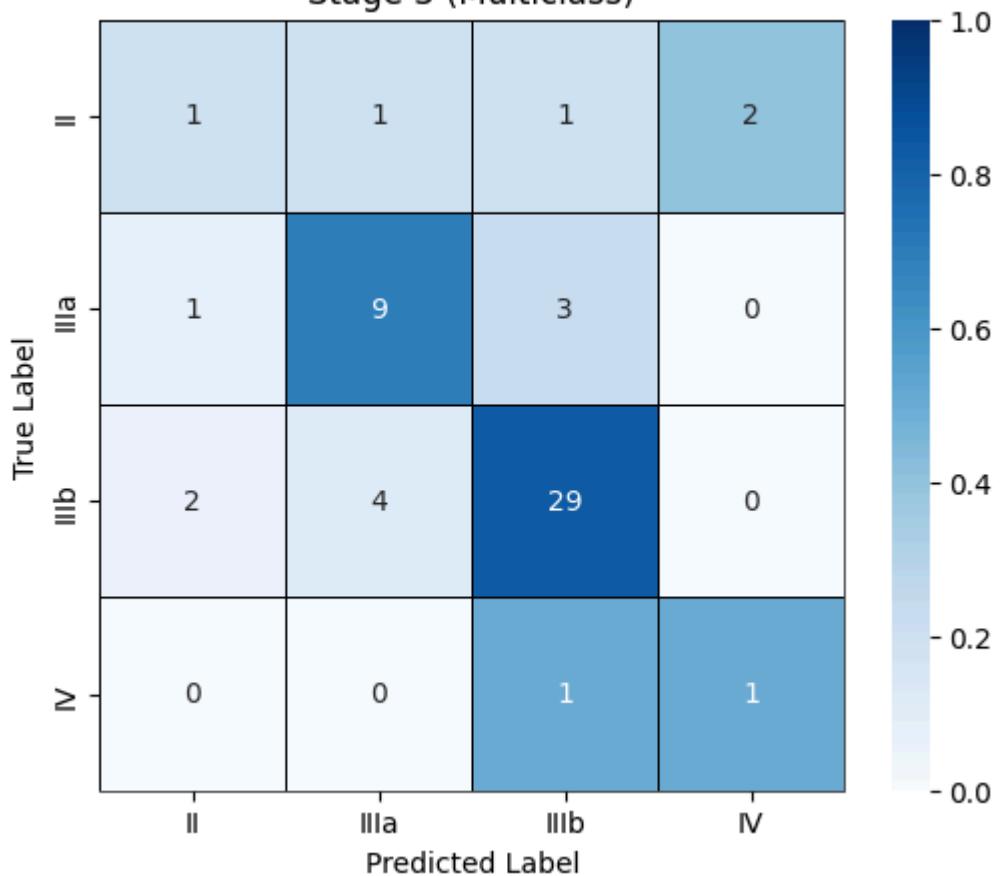
-- Stage 2 (DC vs Rest) --

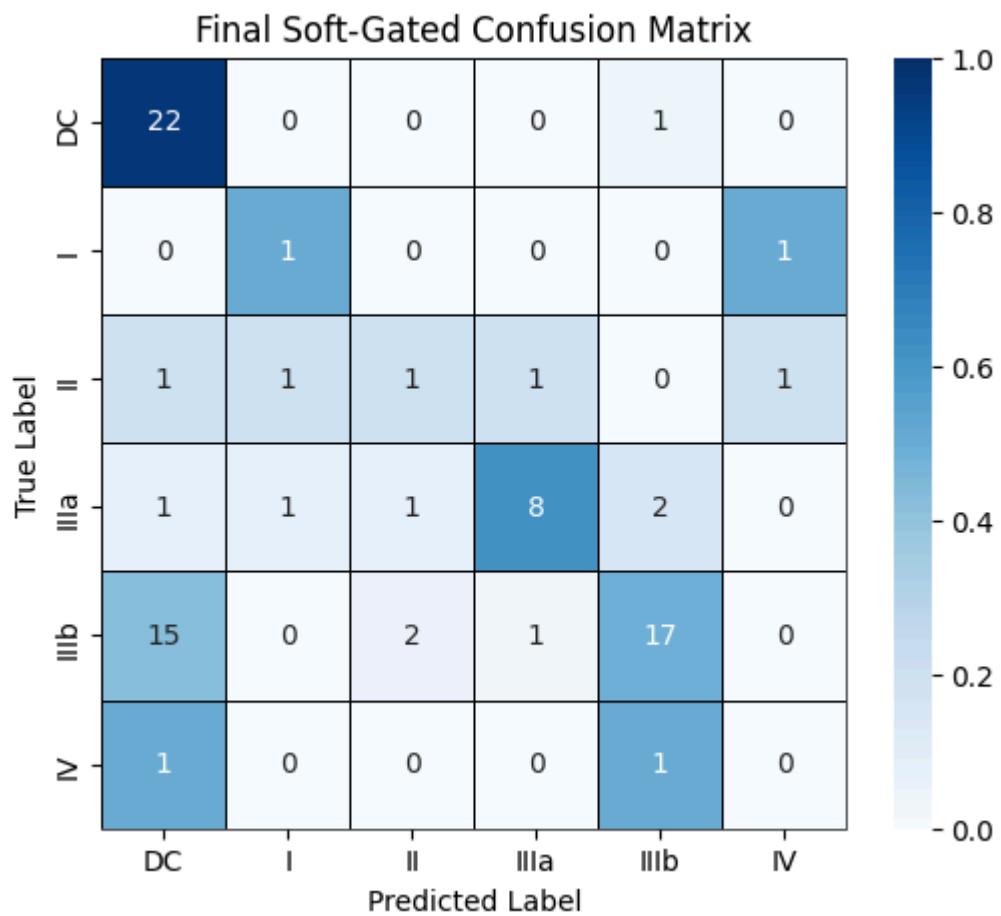
[Threshold Optimization] Best balanced_accuracy: 0.8390 at threshold=0.211



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 2 =====

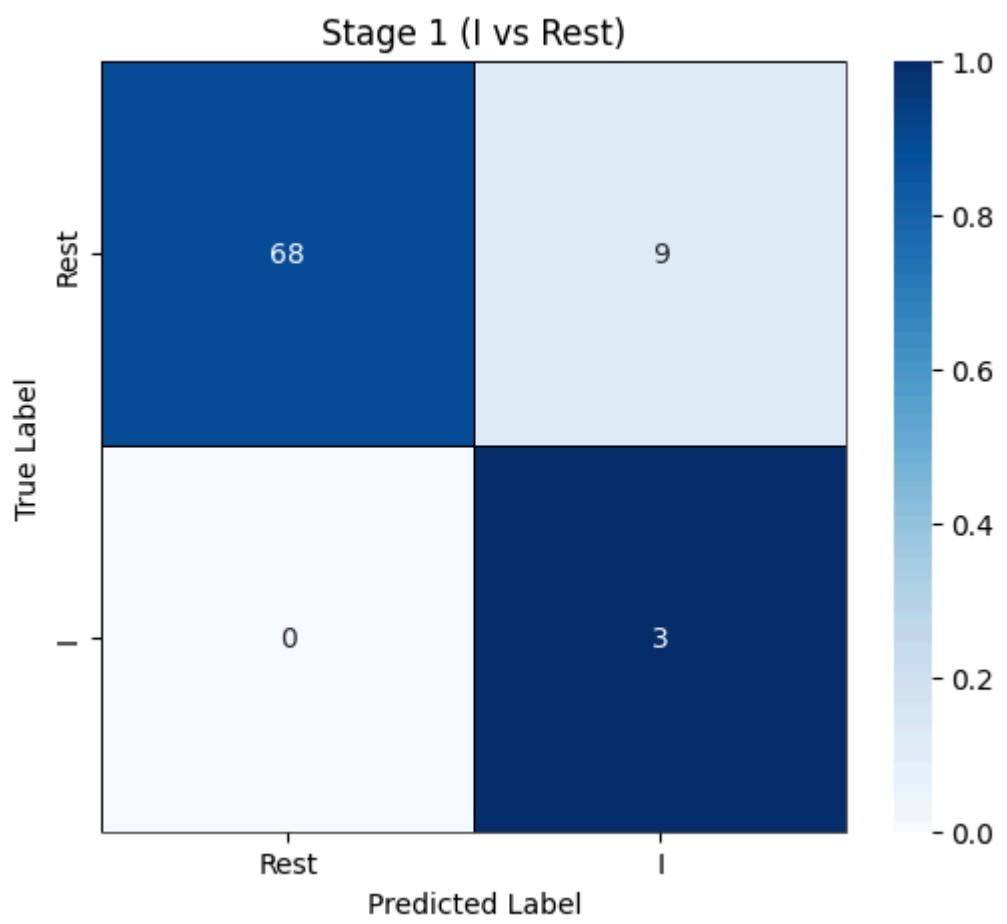
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

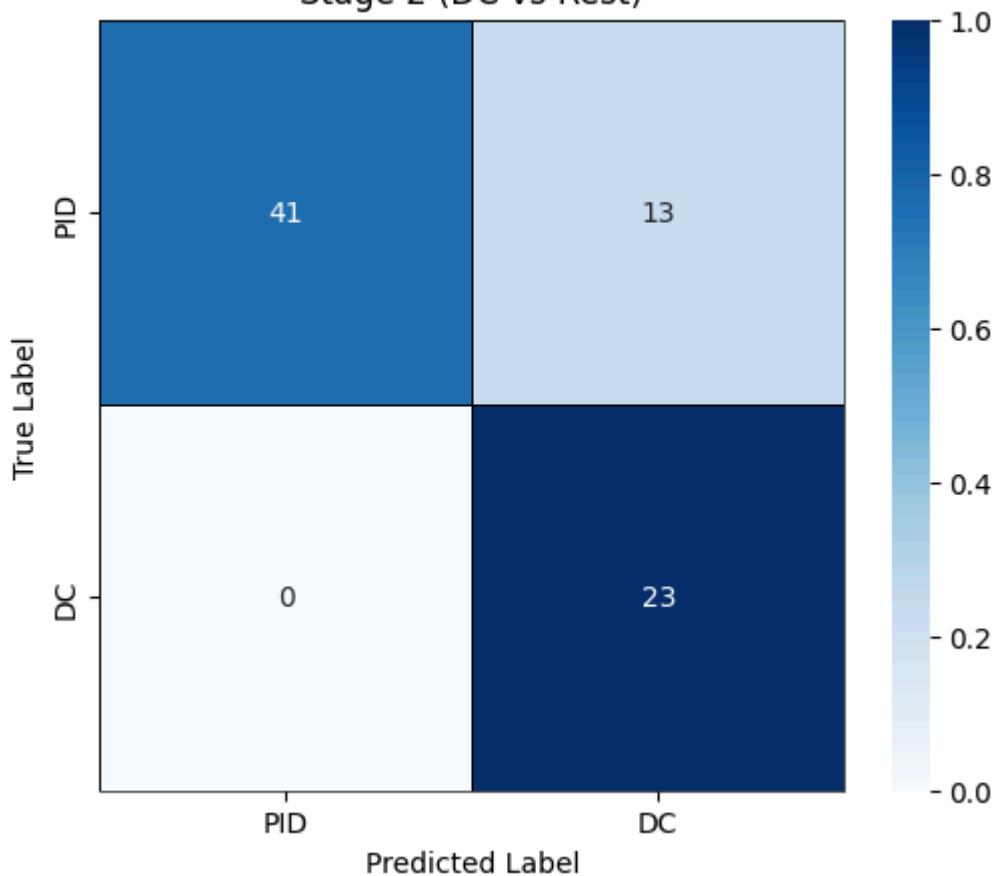
[Threshold Optimization] Best balanced_accuracy: 0.9679 at threshold=0.523



-- Stage 2 (DC vs Rest) --

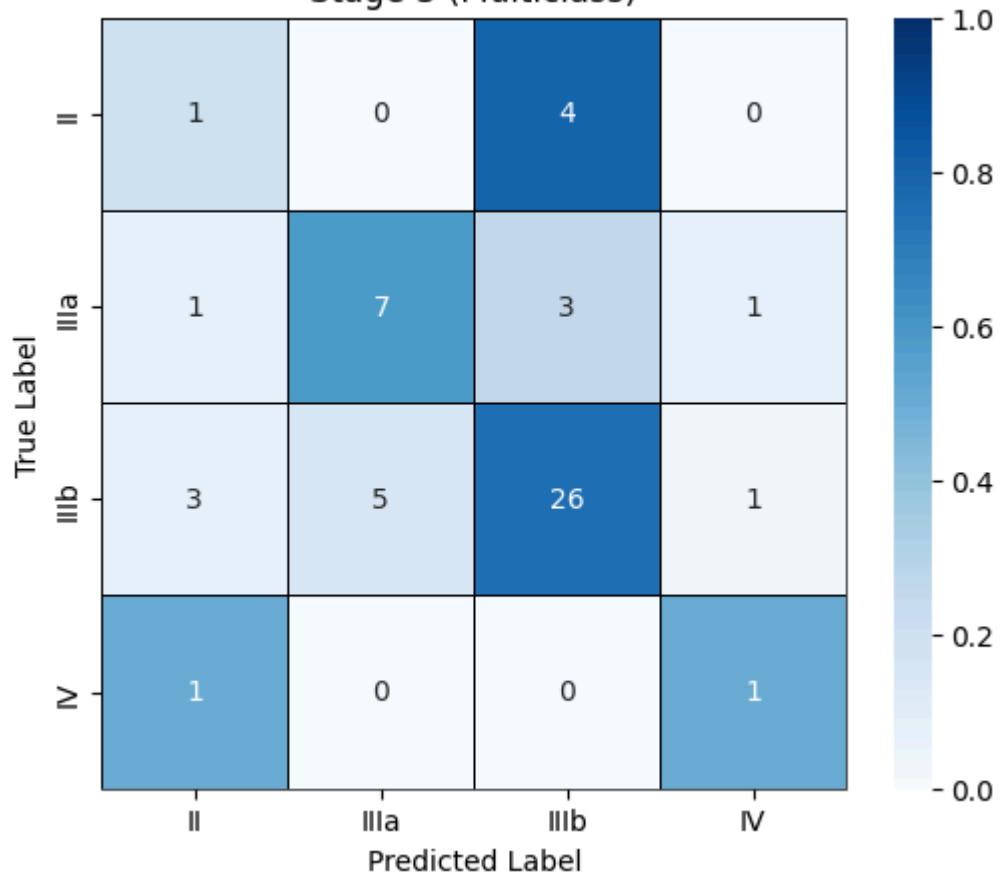
[Threshold Optimization] Best balanced_accuracy: 0.8565 at threshold=0.121

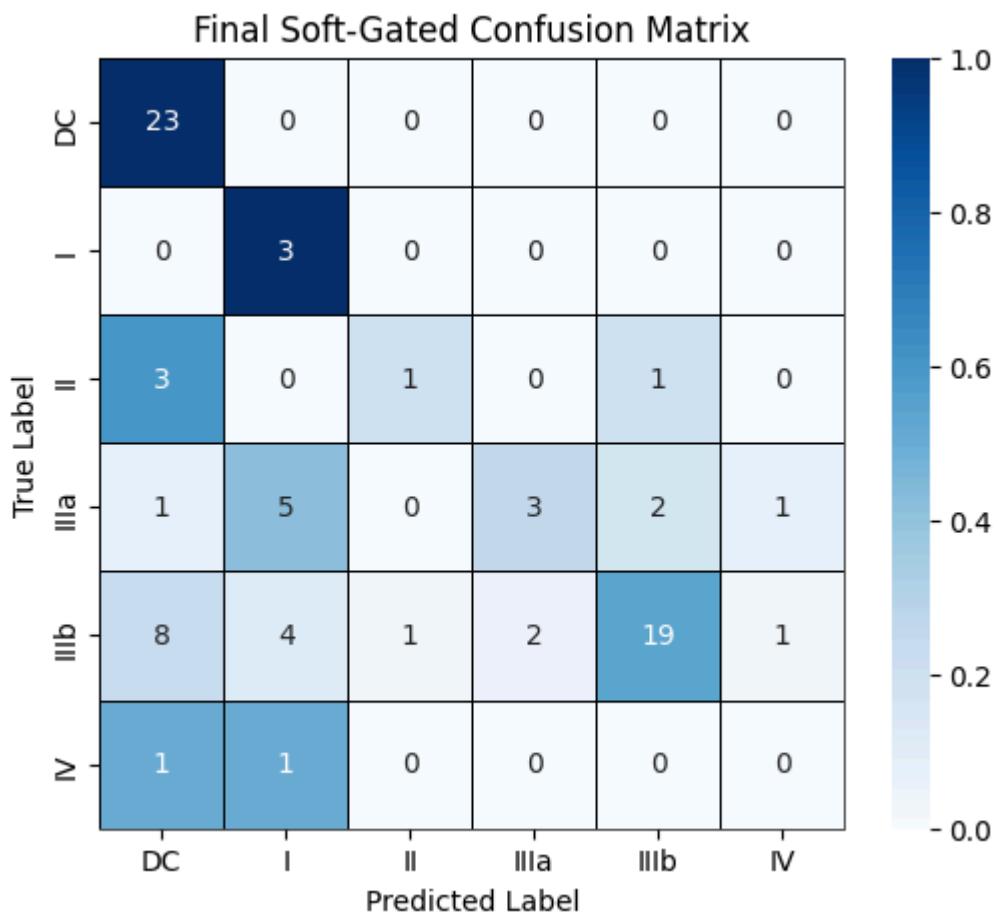
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 3 =====

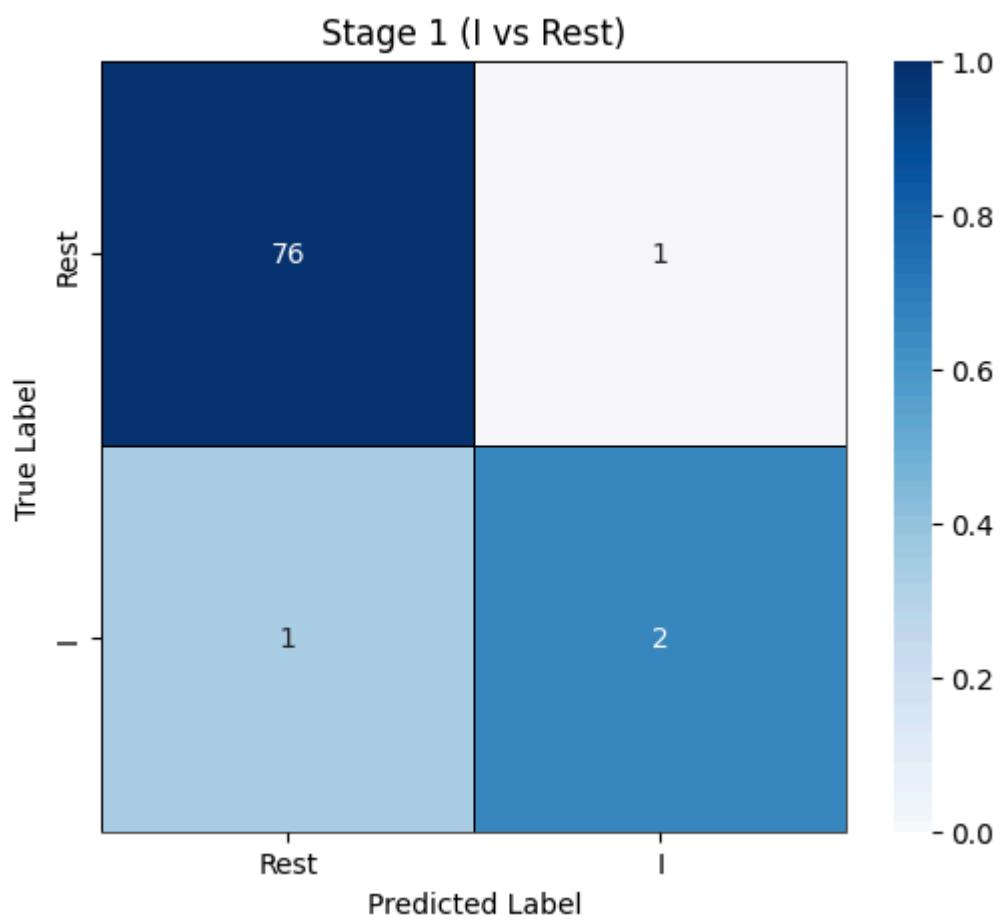
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'MC6. DP T cells', 'C64', 'C100', 'C126', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

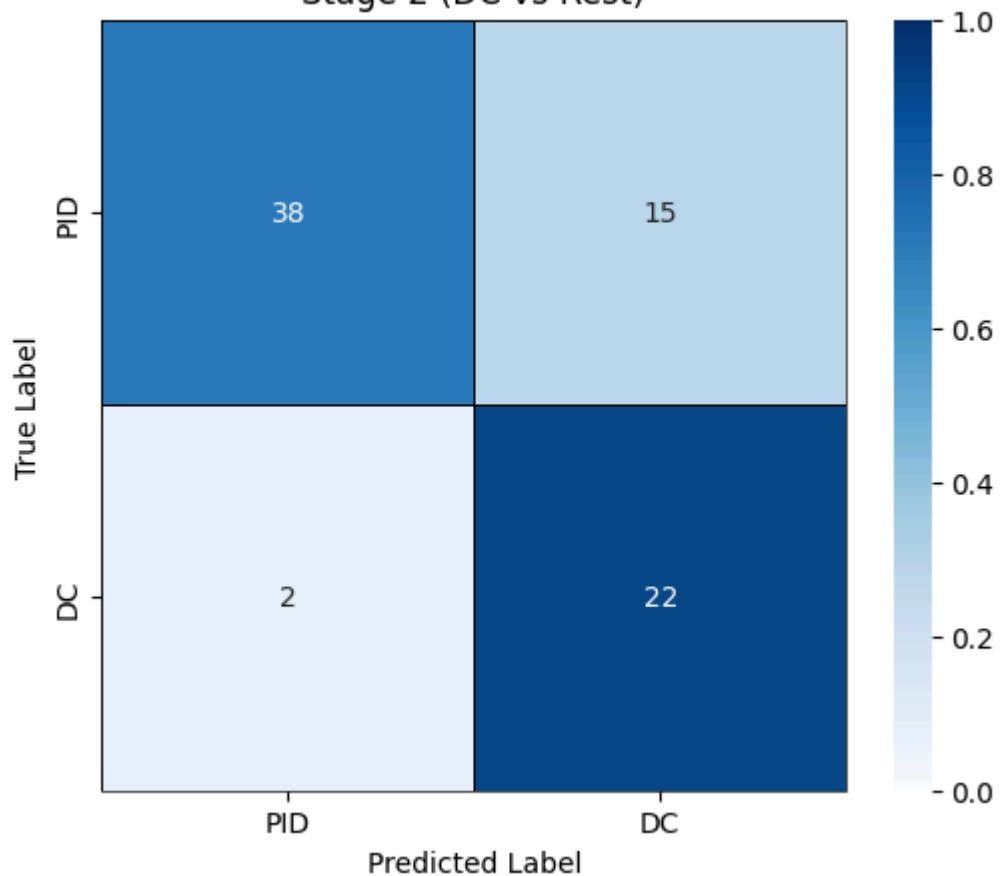
[Threshold Optimization] Best balanced_accuracy: 0.9936 at threshold=0.407



-- Stage 2 (DC vs Rest) --

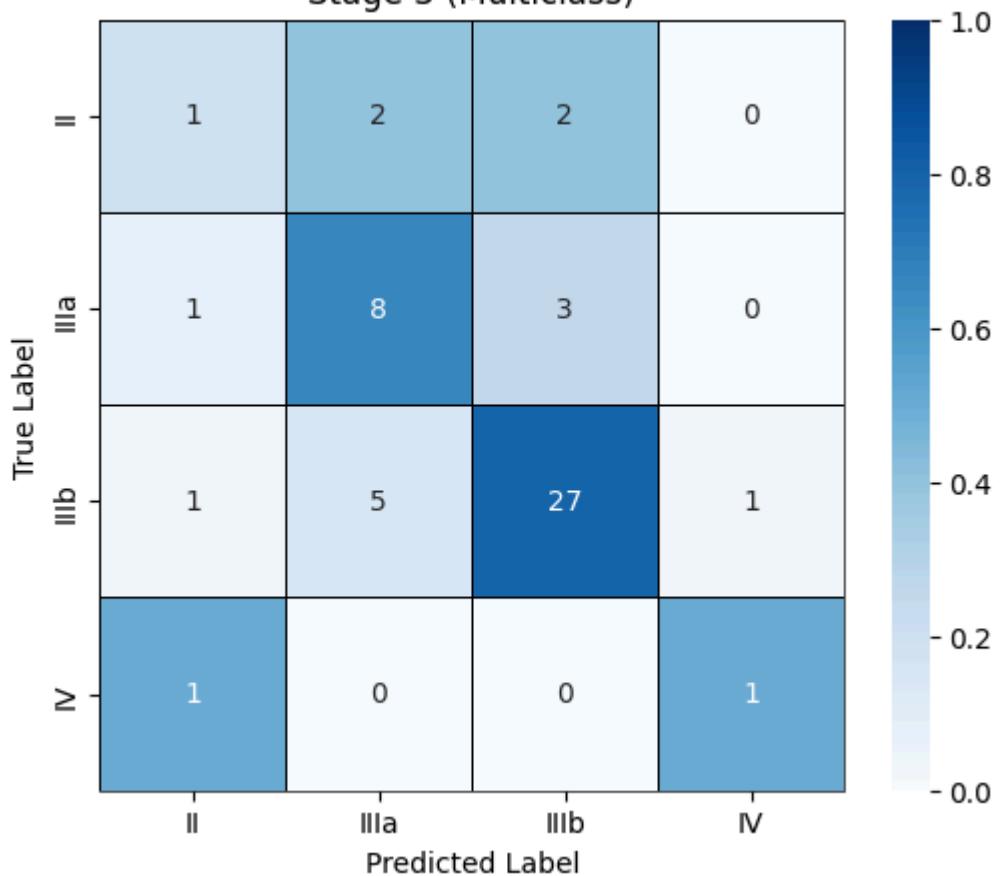
[Threshold Optimization] Best balanced_accuracy: 0.8146 at threshold=0.080

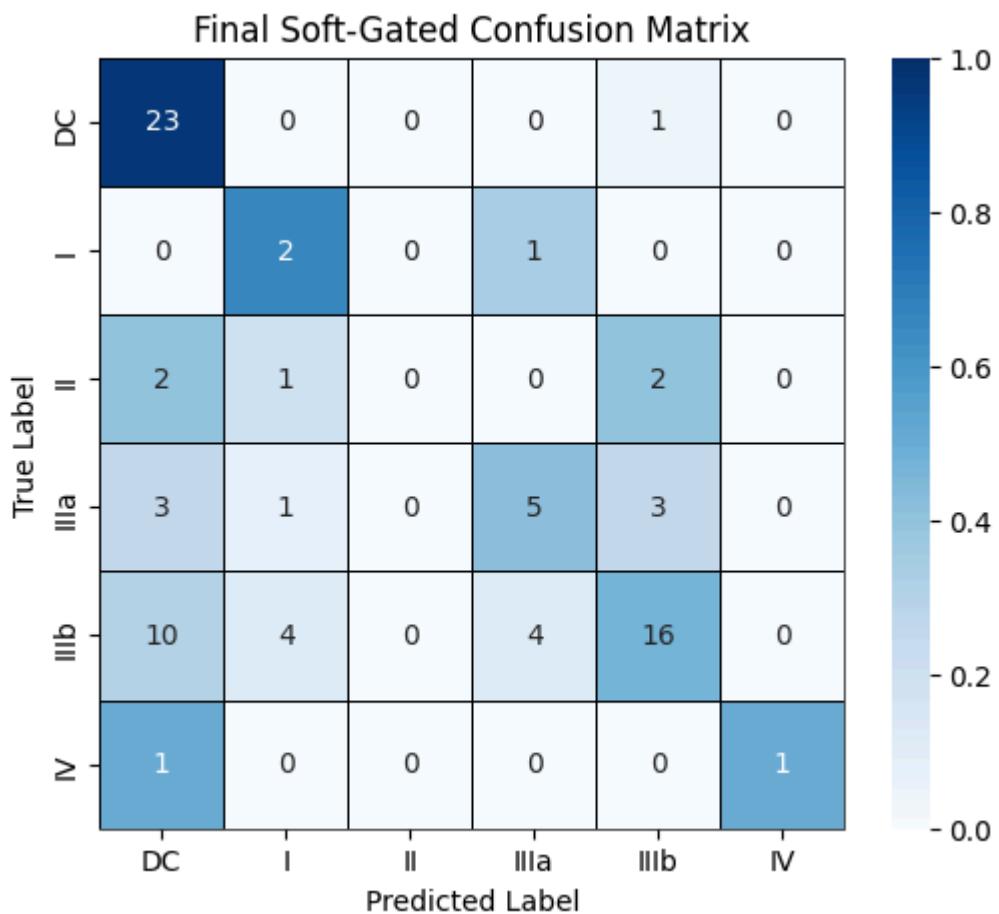
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 4 =====

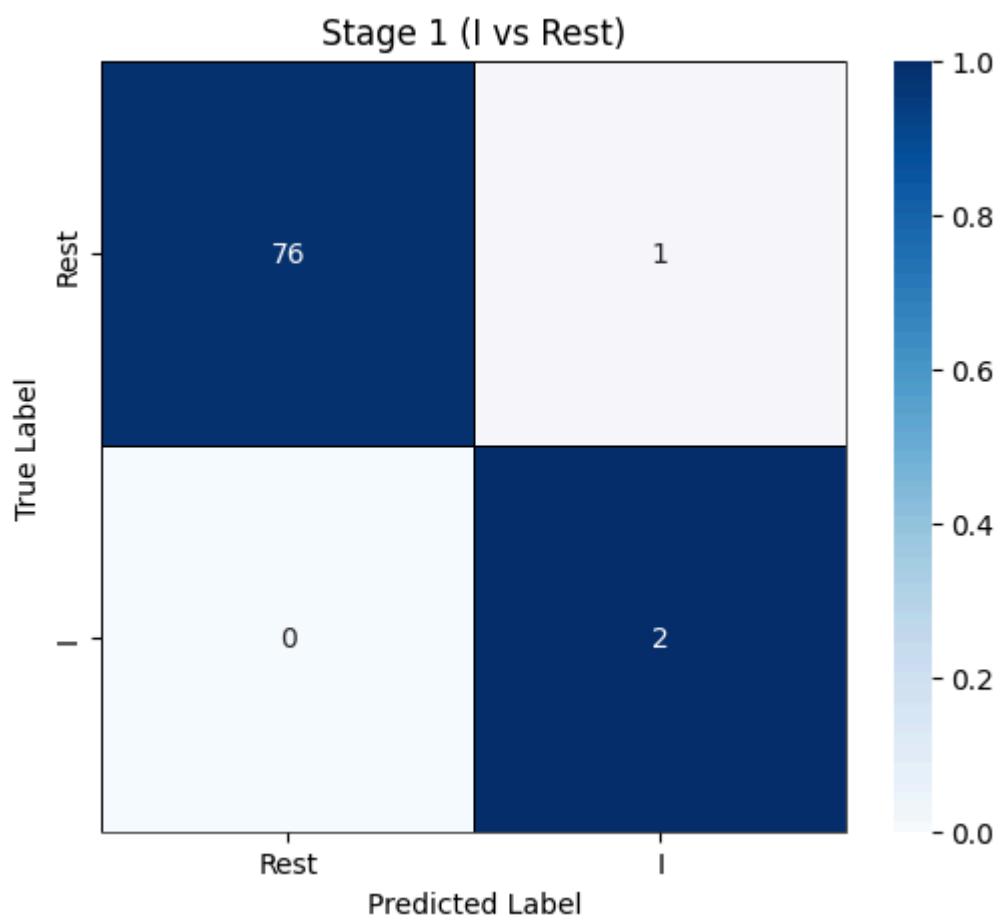
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

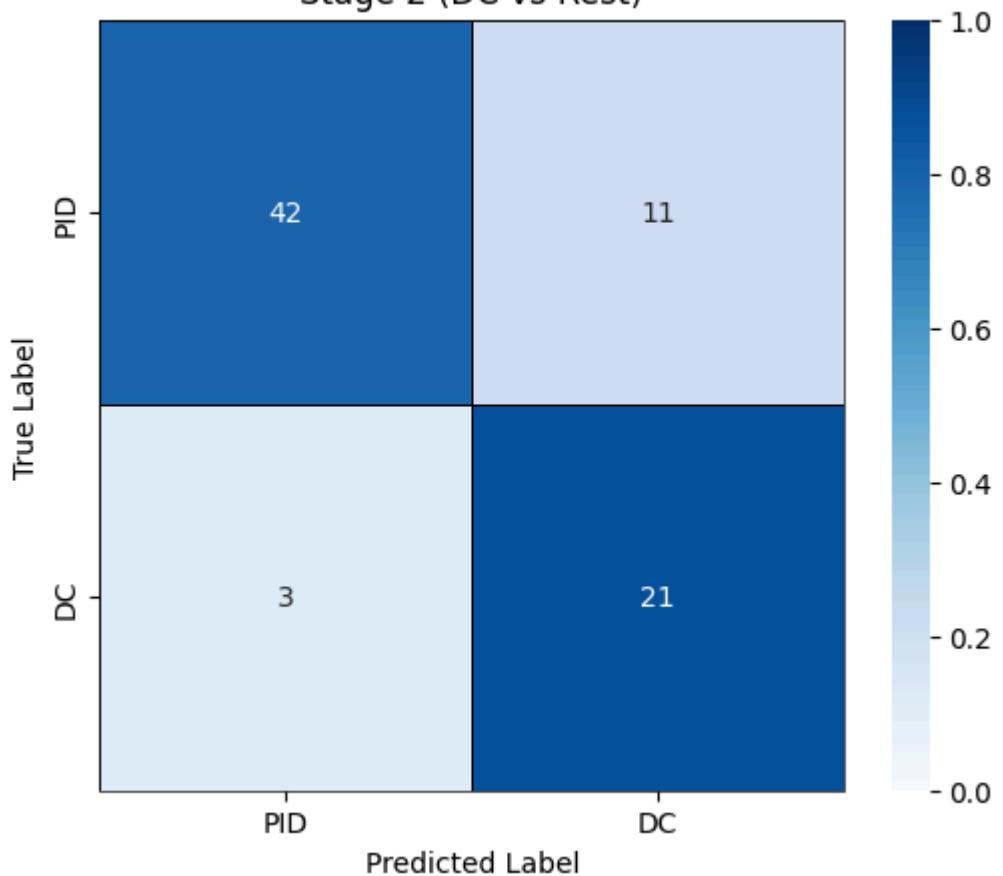
[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.492



-- Stage 2 (DC vs Rest) --

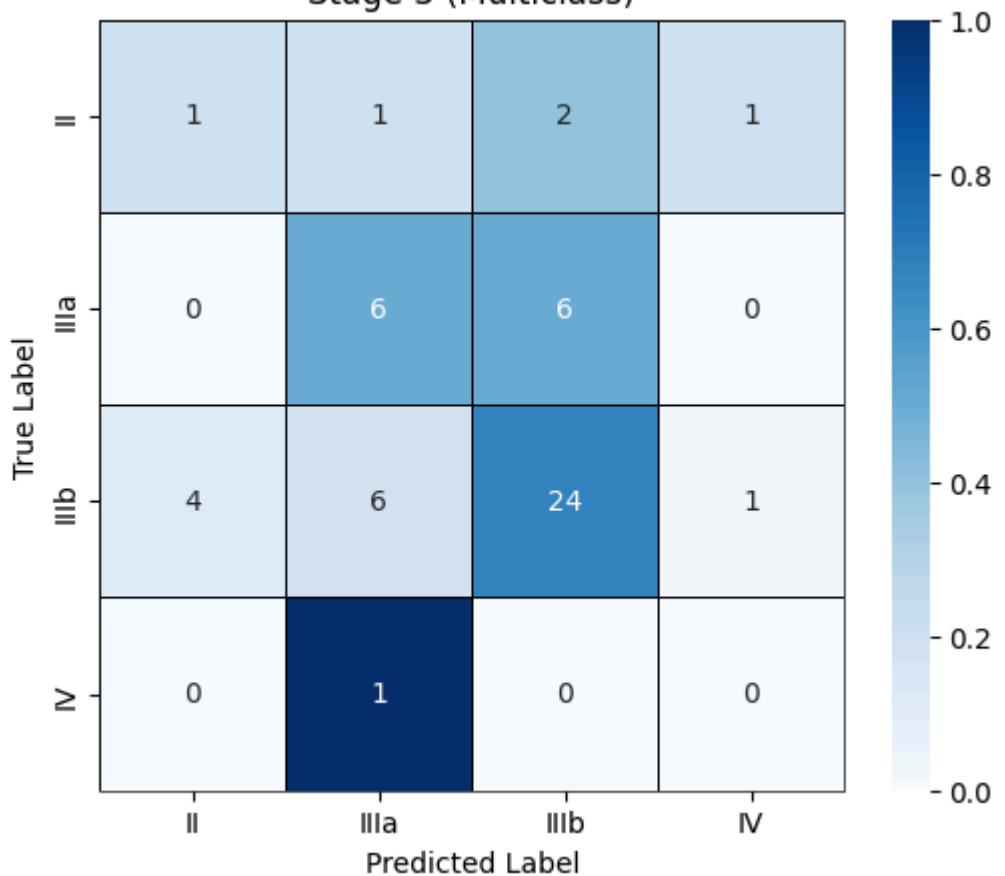
[Threshold Optimization] Best balanced_accuracy: 0.8383 at threshold=0.186

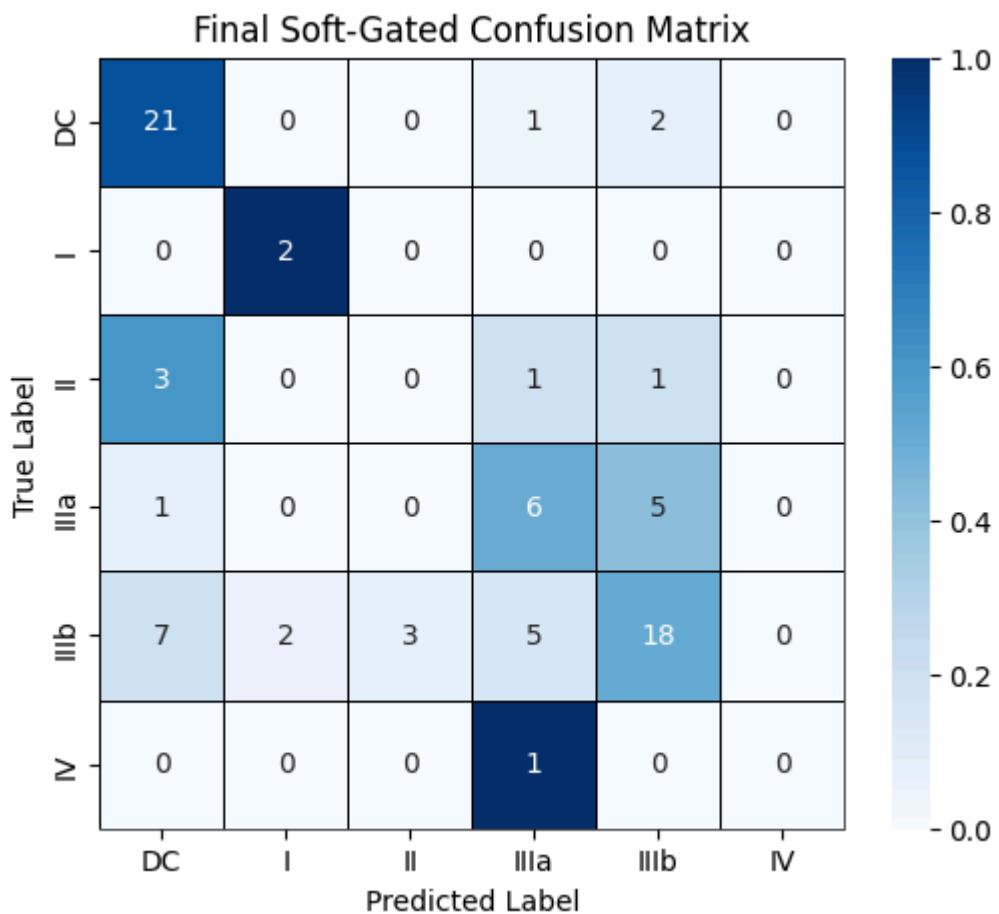
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 5 =====

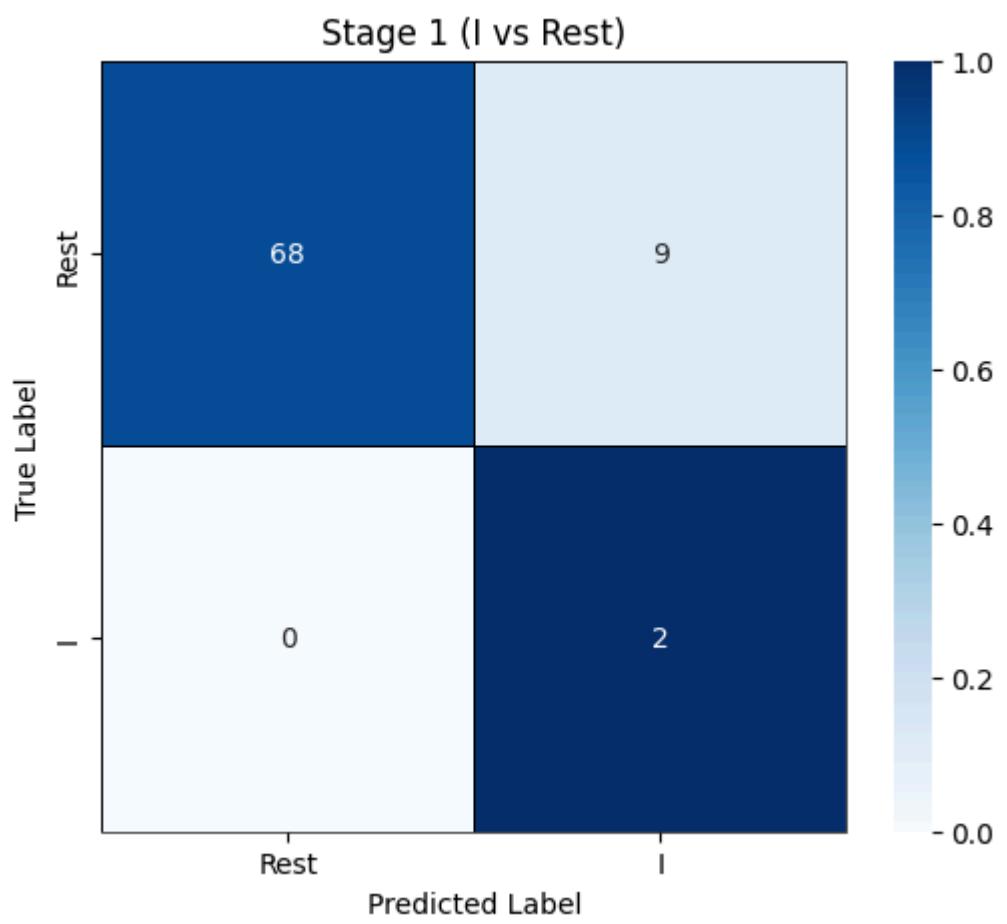
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']`

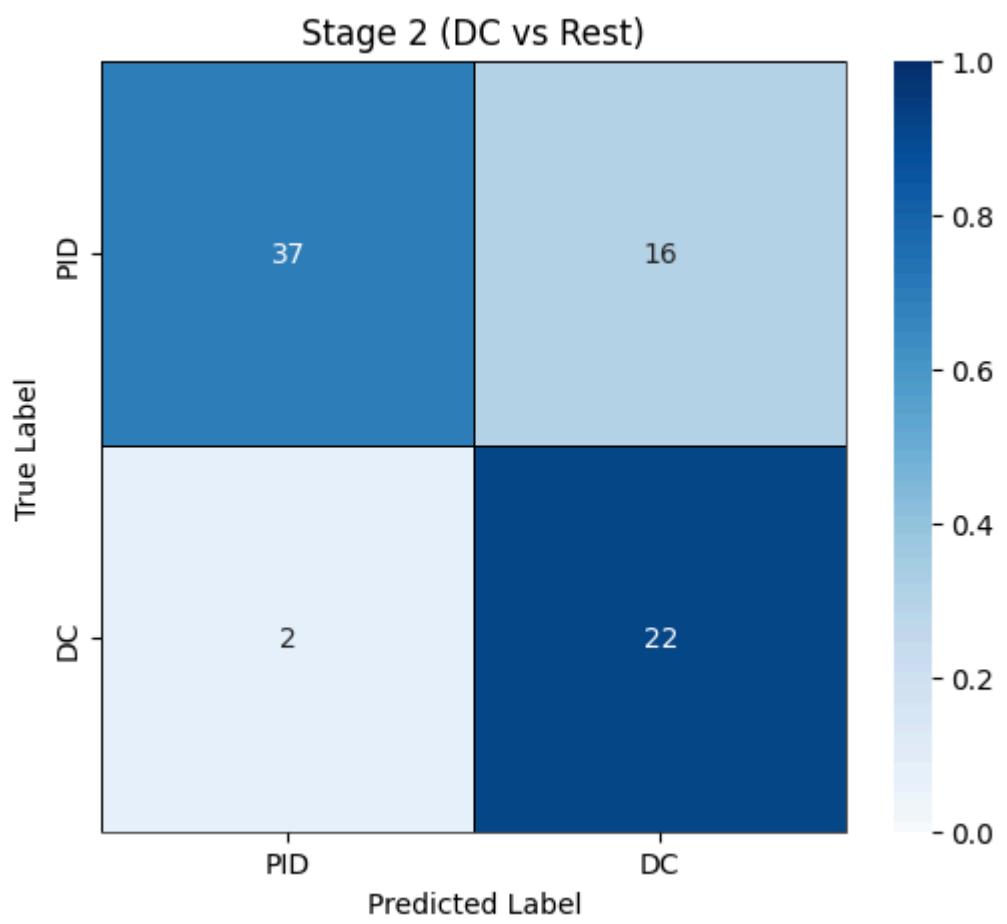
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 0.9744 at threshold=0.593



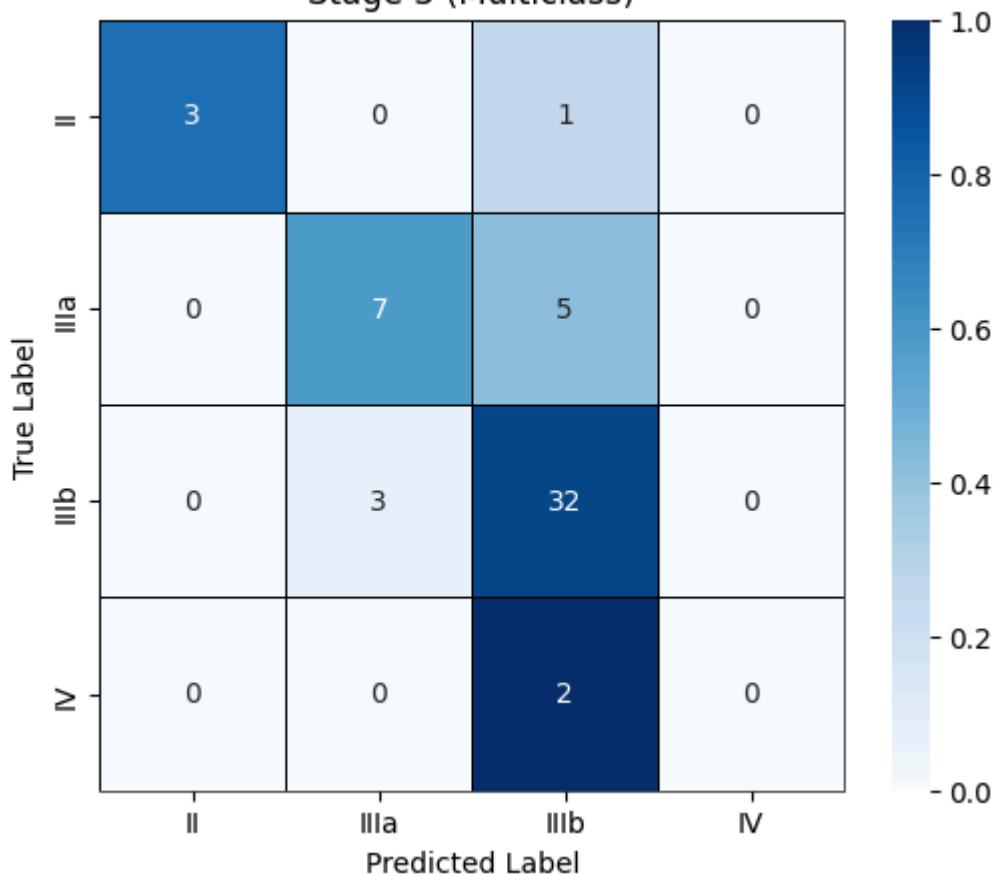
-- Stage 2 (DC vs Rest) --

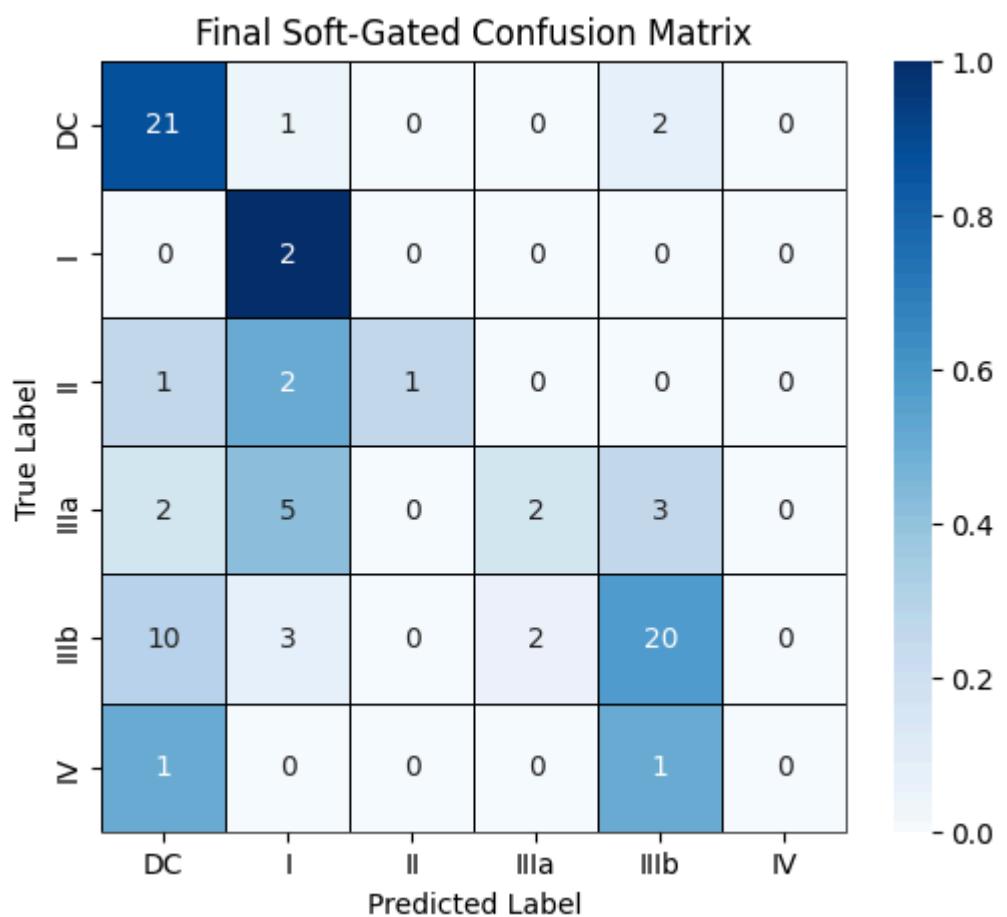
[Threshold Optimization] Best balanced_accuracy: 0.8747 at threshold=0.151



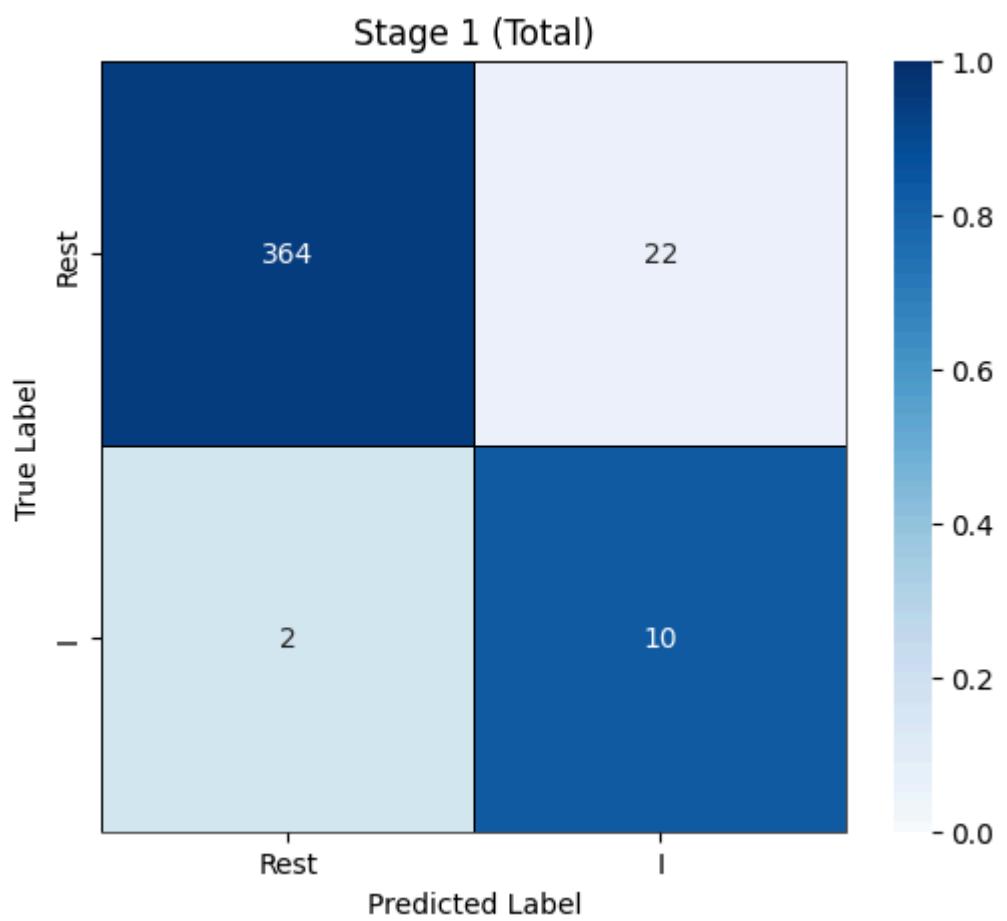
-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)

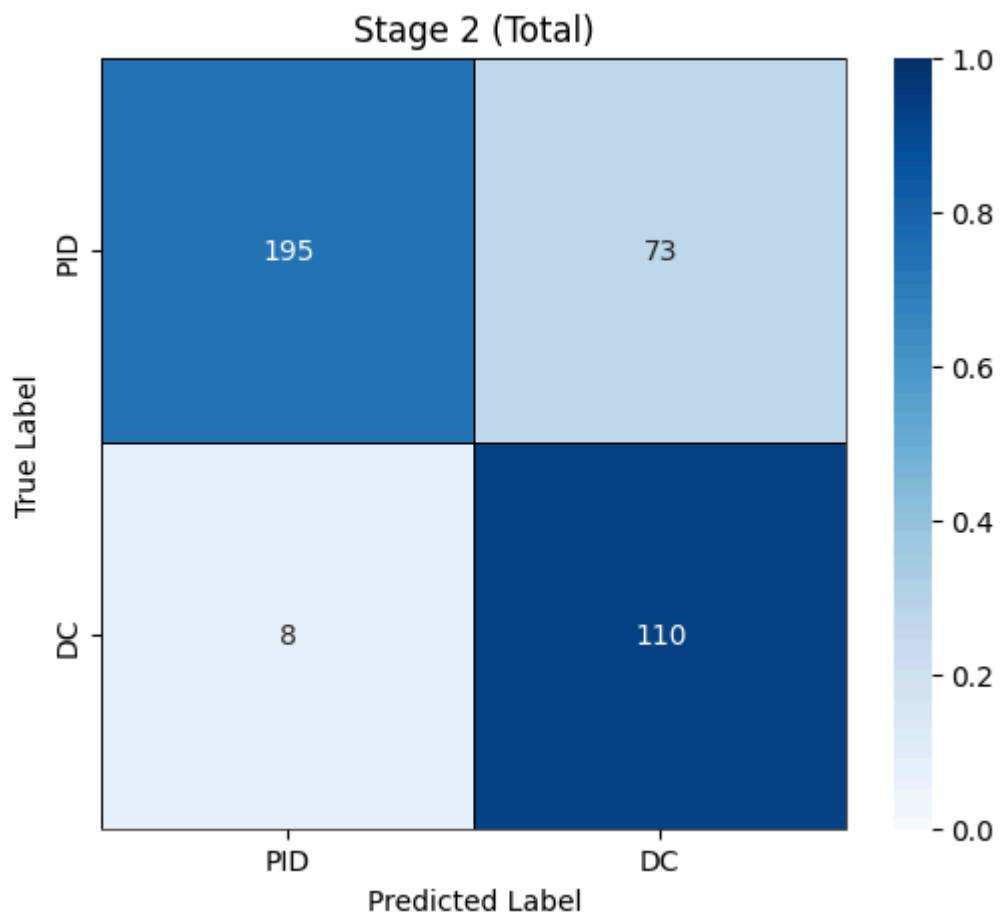




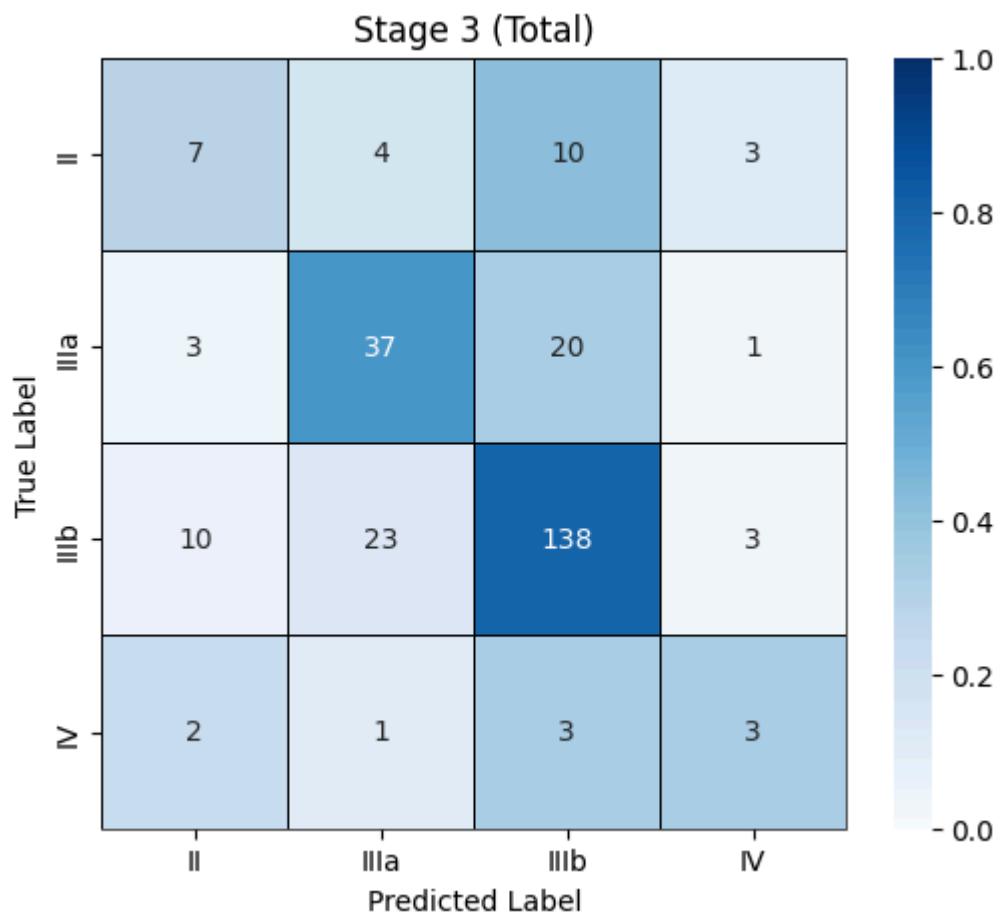
===== AGGREGATED CONFUSION MATRICES =====



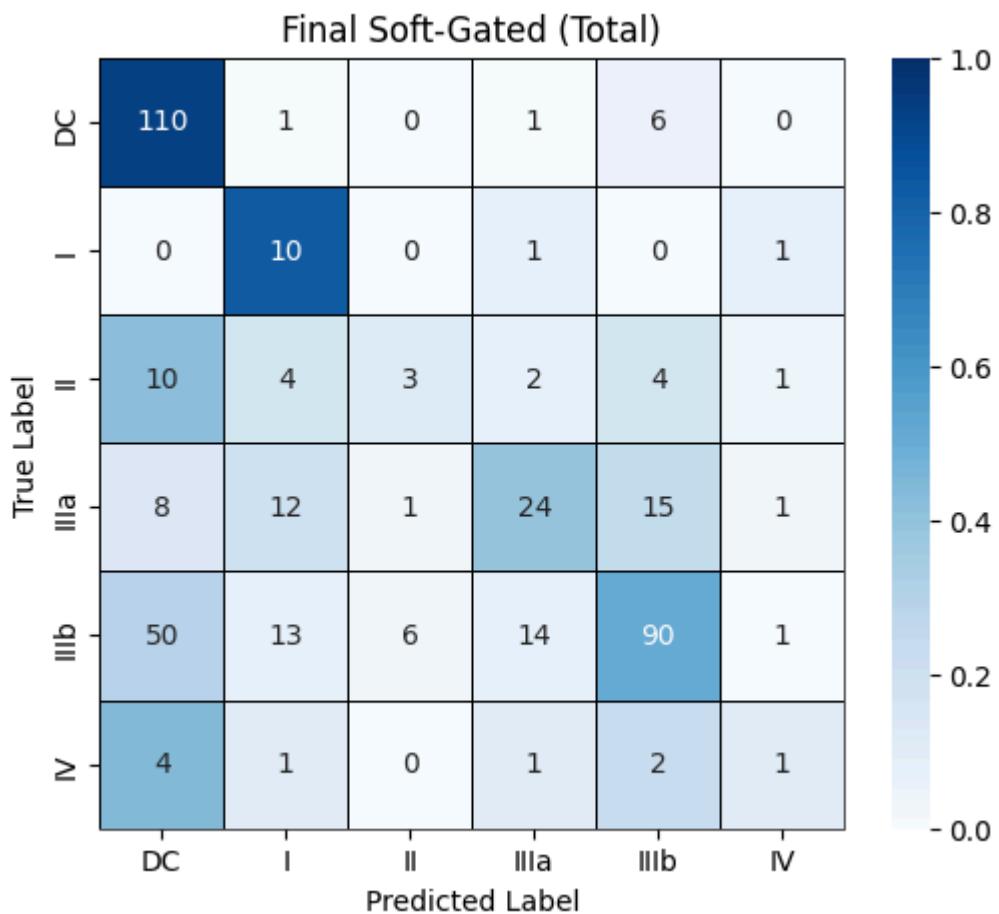
	precision	recall	f1-score	support
0	0.995	0.943	0.968	386
1	0.312	0.833	0.455	12
accuracy		0.940	0.940	398
macro avg	0.654	0.888	0.711	398
weighted avg	0.974	0.940	0.953	398



	precision	recall	f1-score	support
0	0.961	0.728	0.828	268
1	0.601	0.932	0.731	118
accuracy		0.790	0.790	386
macro avg	0.781	0.830	0.779	386
weighted avg	0.851	0.790	0.798	386



	precision	recall	f1-score	support
0	0.318	0.292	0.304	24
1	0.569	0.607	0.587	61
2	0.807	0.793	0.800	174
3	0.300	0.333	0.316	9
accuracy		0.690	0.690	268
macro avg	0.499	0.506	0.502	268
weighted avg	0.692	0.690	0.691	268



===== CV SUMMARY =====

precision recall f1-score support

DC	0.6044	0.9322	0.7333	118
I	0.2439	0.8333	0.3774	12
II	0.3000	0.1250	0.1765	24
IIIa	0.5581	0.3934	0.4615	61
IIIb	0.7692	0.5172	0.6186	174
IV	0.2000	0.1111	0.1429	9

accuracy		0.5980	398	
macro avg	0.4459	0.4854	0.4184	398
weighted avg	0.6310	0.5980	0.5838	398

Balanced Accuracy (Final Soft-Gated, aggregated): 0.48538863942268623

===== Accuracies =====

Per-fold balanced accuracy:

Stage 1: [0.7371794871794872, 0.9415584415584415, 0.8268398268398268, 0.9935064935064934, 0.9415584415584415]

Stage 2: [0.8146245059288537, 0.8796296296296297, 0.8168238993710691, 0.8337264150943396, 0.8073899371069182]

```
Stage 3: [0.5552197802197802, 0.506547619047619, 0.5401960784313725,  
0.3464285714285714, 0.5619047619047619]  
Mean per-fold balanced accuracy: [np.float64(0.888128538128538),  
np.float64(0.8304388774261622), np.float64(0.502059362206421)]  
Aggregated balanced accuracy:  
Stage 1: 0.8881692573402418  
Stage 2: 0.829907665064508  
Stage 3: 0.5061652063312606  
Final Soft-Gated: 0.48538863942268623
```

```
===== CV RUN 3/5 =====  
Starting 5-split CV (3 train / 1 val / 1 test) on 398 samples...
```

```
===== Fold 1 =====
```

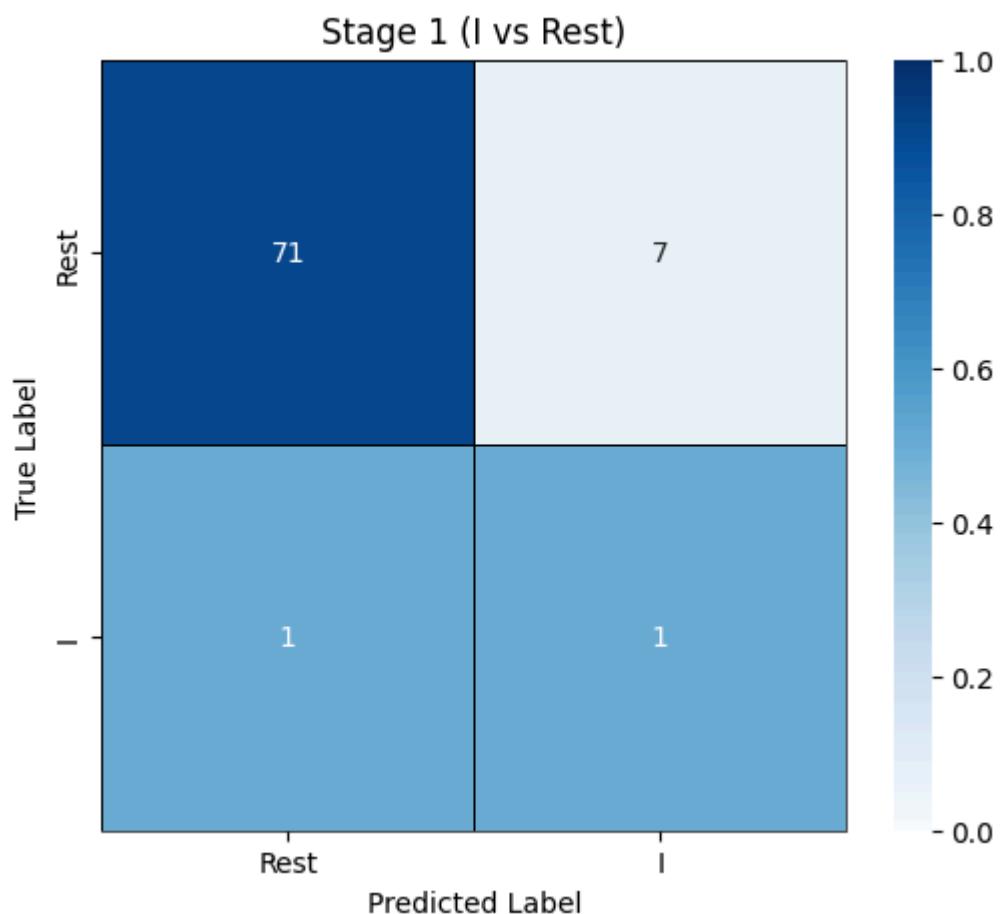
```
Train: 238, Val: 80, Test: 80
```

```
These features will be dropped:
```

```
['C11', 'C101', 'C64', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86',  
'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']
```

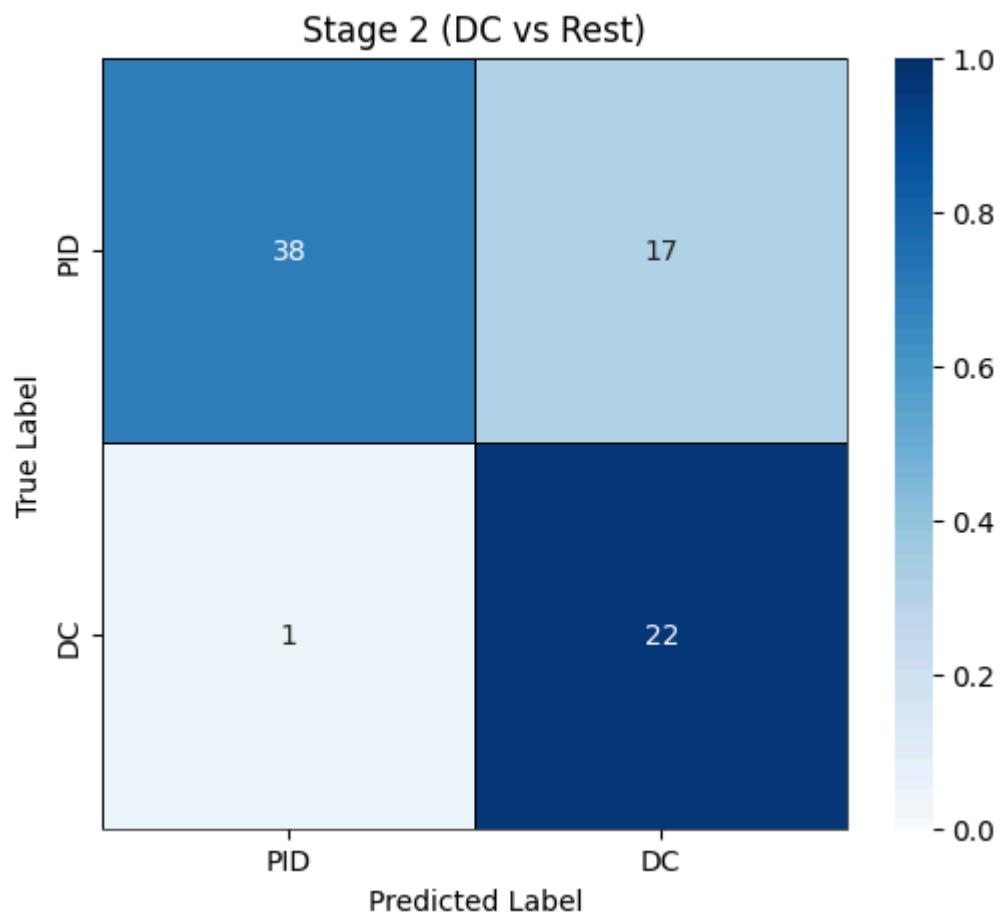
```
-- Stage 1 (I vs Rest) --
```

```
[Threshold Optimization] Best balanced_accuracy: 0.9935 at threshold=0.638
```



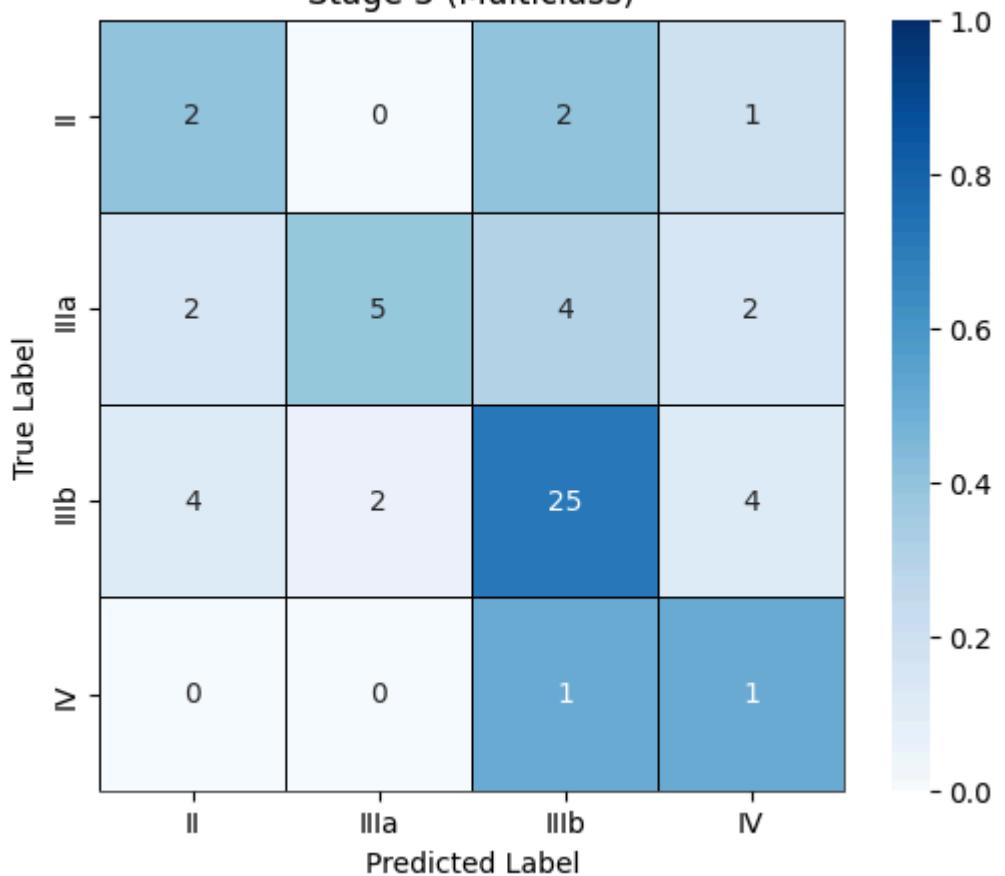
-- Stage 2 (DC vs Rest) --

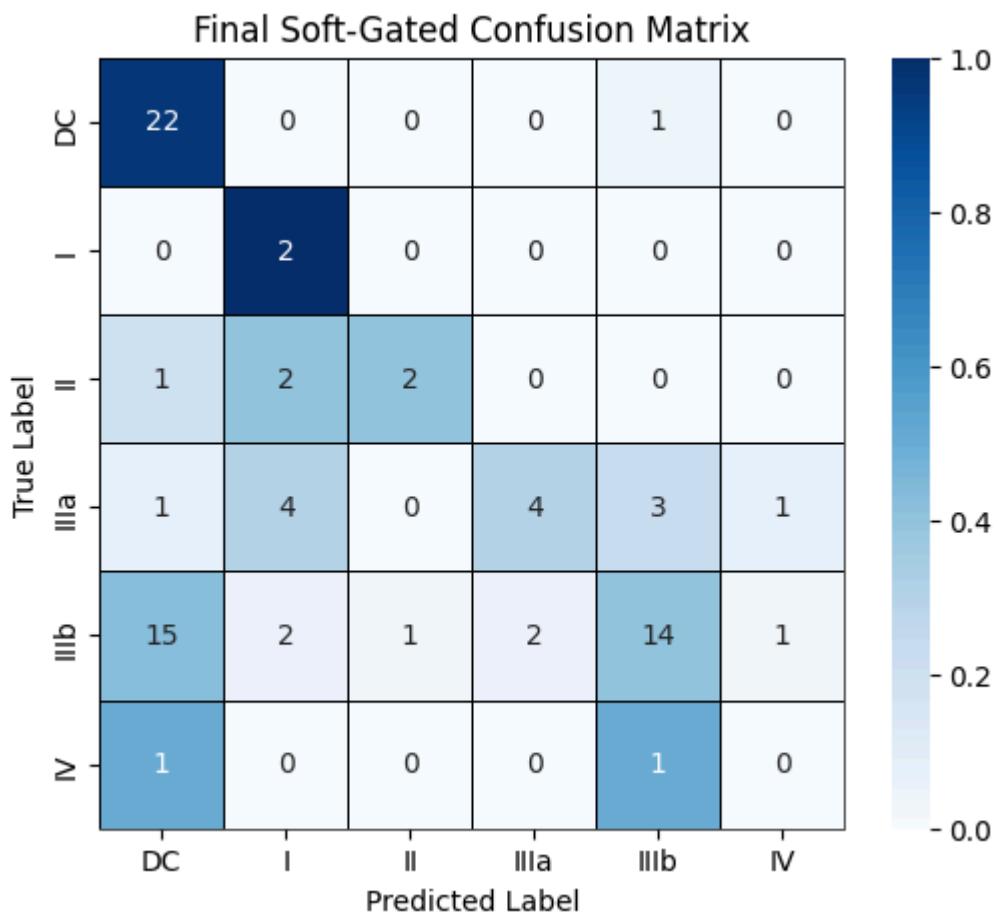
[Threshold Optimization] Best balanced_accuracy: 0.8482 at threshold=0.191



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 2 =====

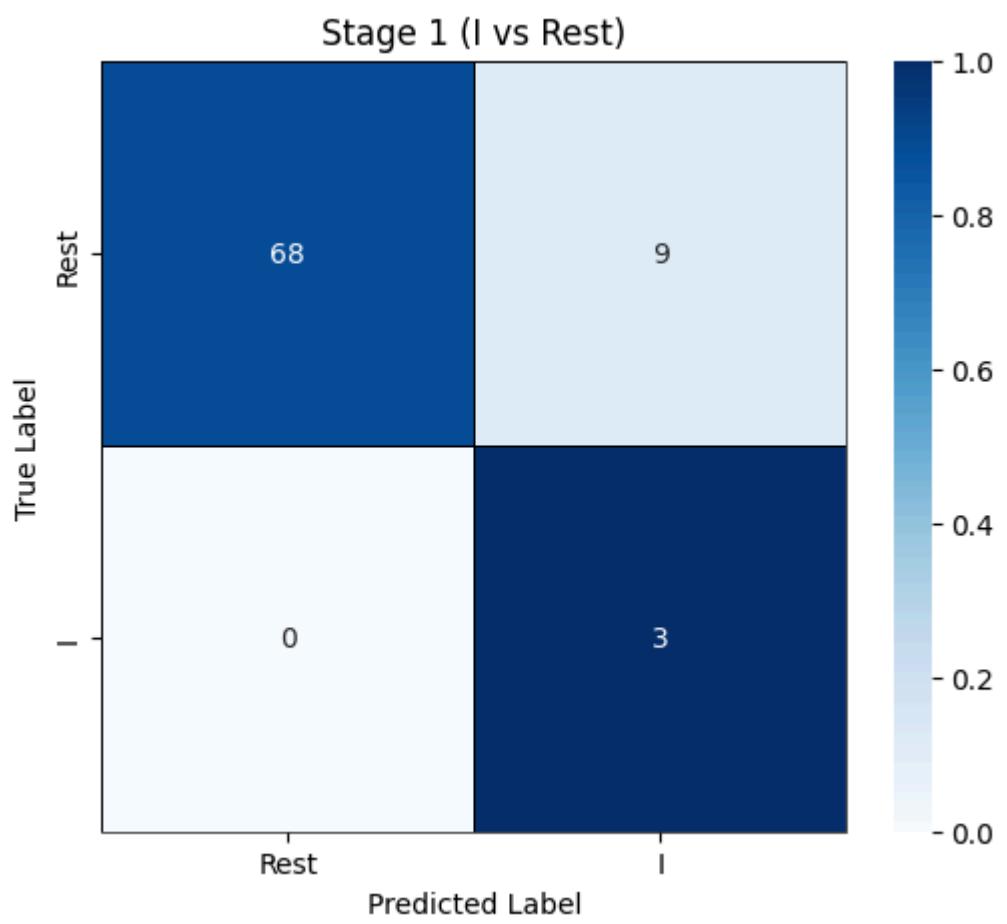
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

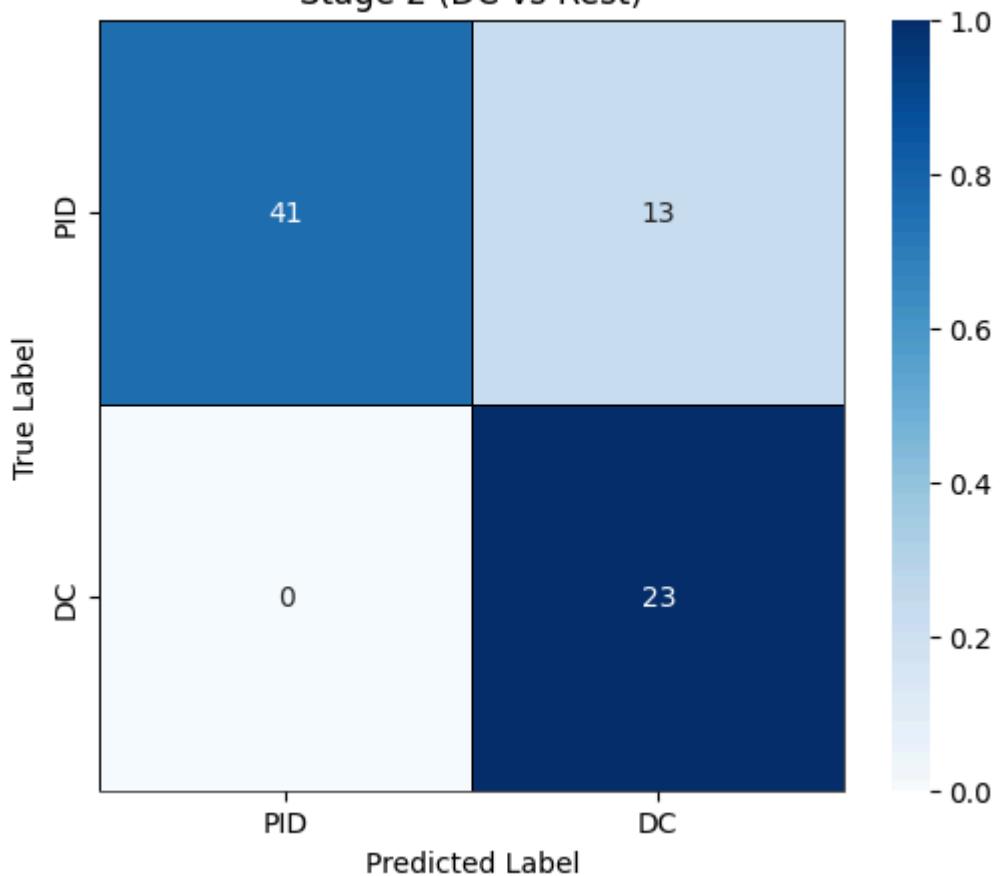
[Threshold Optimization] Best balanced_accuracy: 0.9872 at threshold=0.633



-- Stage 2 (DC vs Rest) --

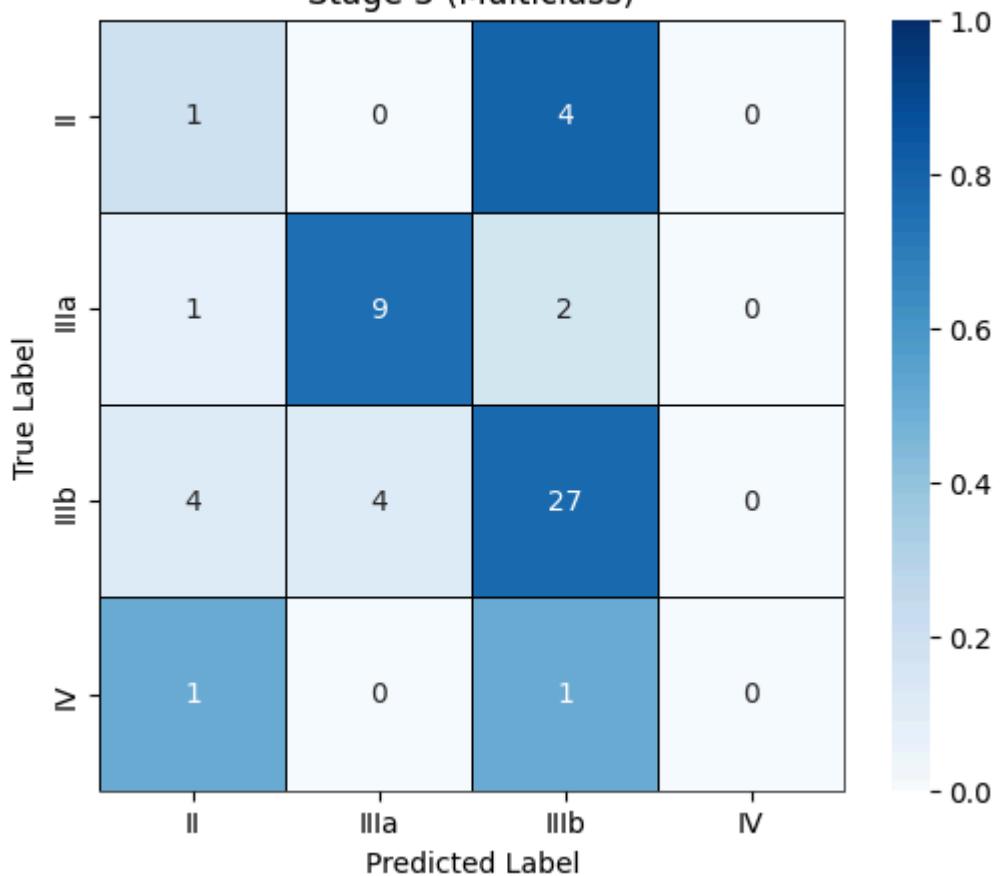
[Threshold Optimization] Best balanced_accuracy: 0.8530 at threshold=0.281

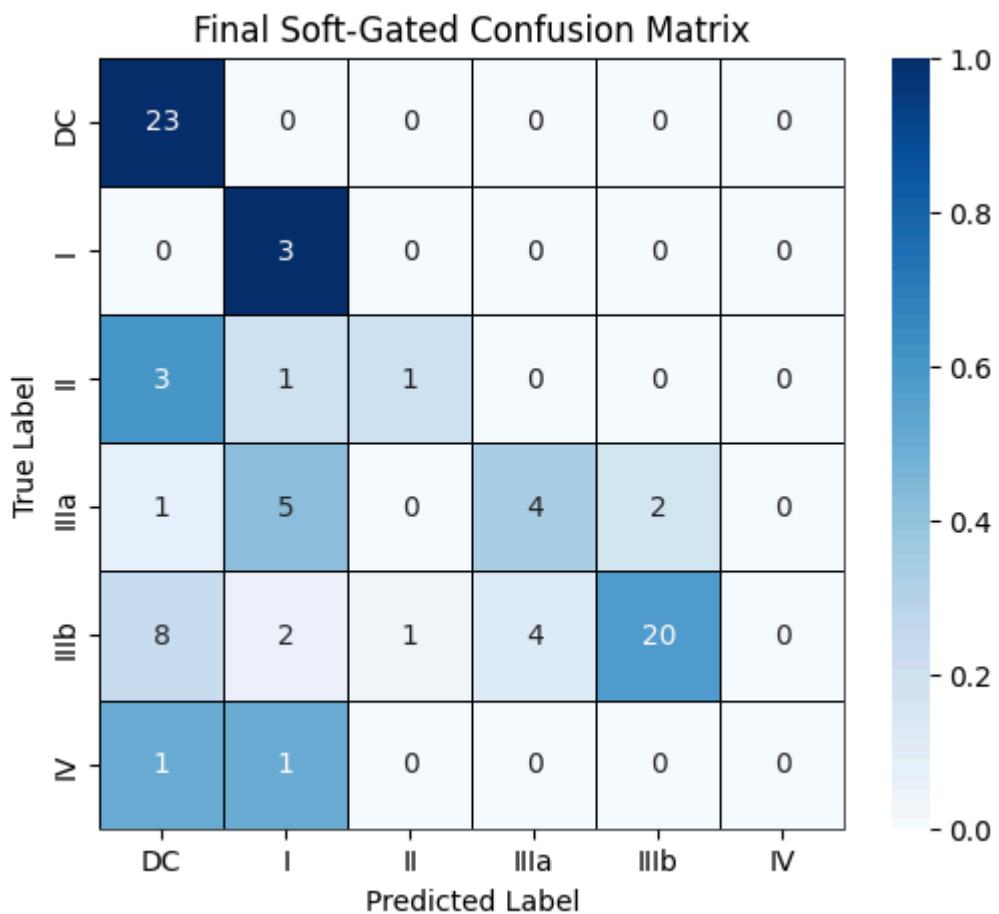
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 3 =====

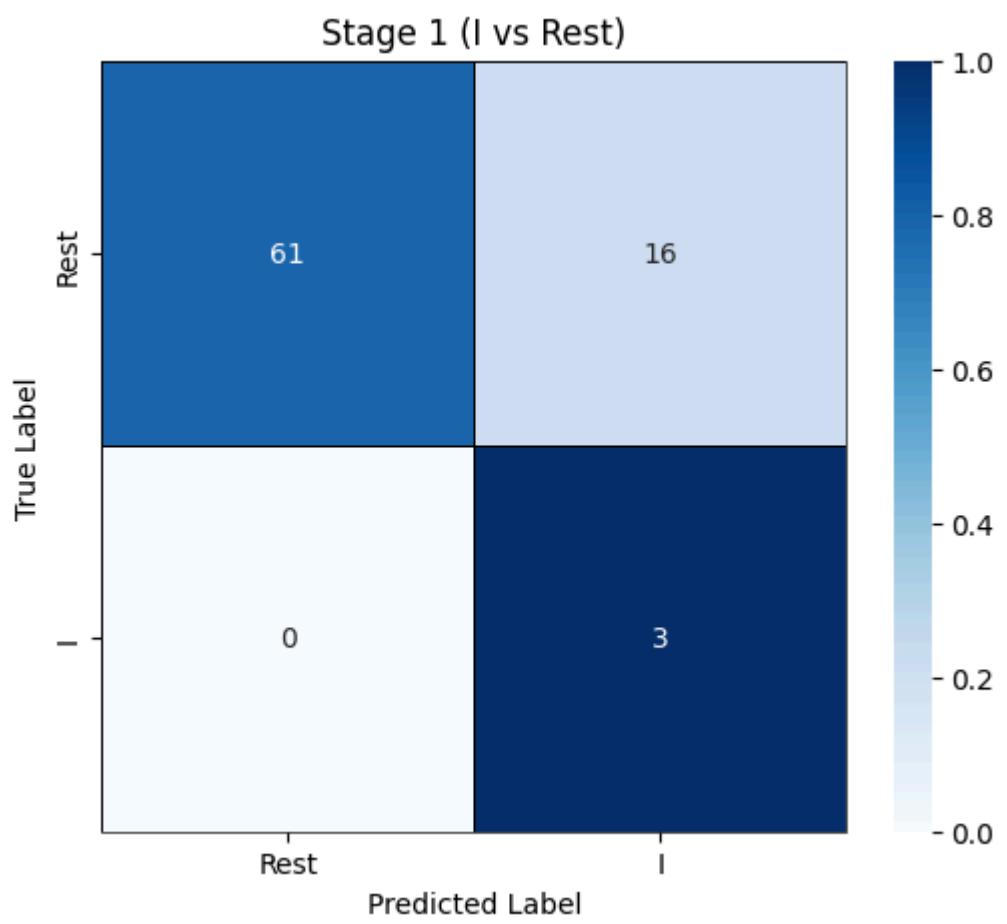
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'MC6. DP T cells', 'C64', 'C100', 'C126', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

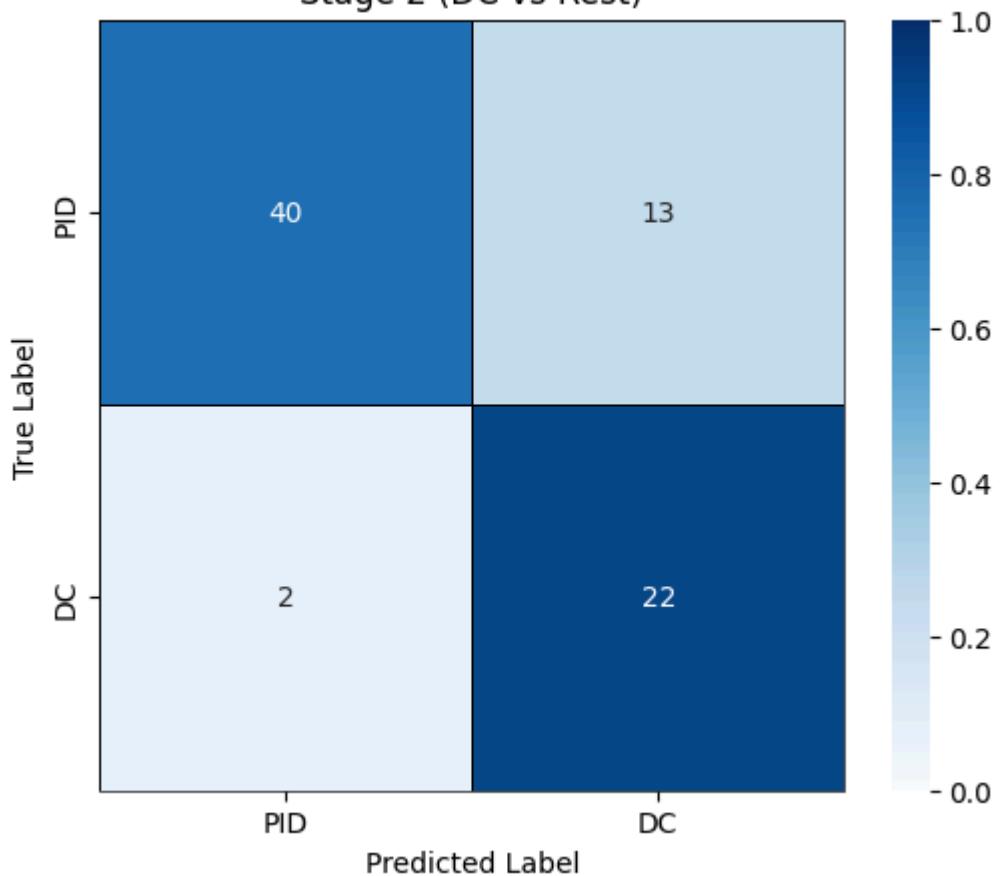
[Threshold Optimization] Best balanced_accuracy: 0.9808 at threshold=0.427



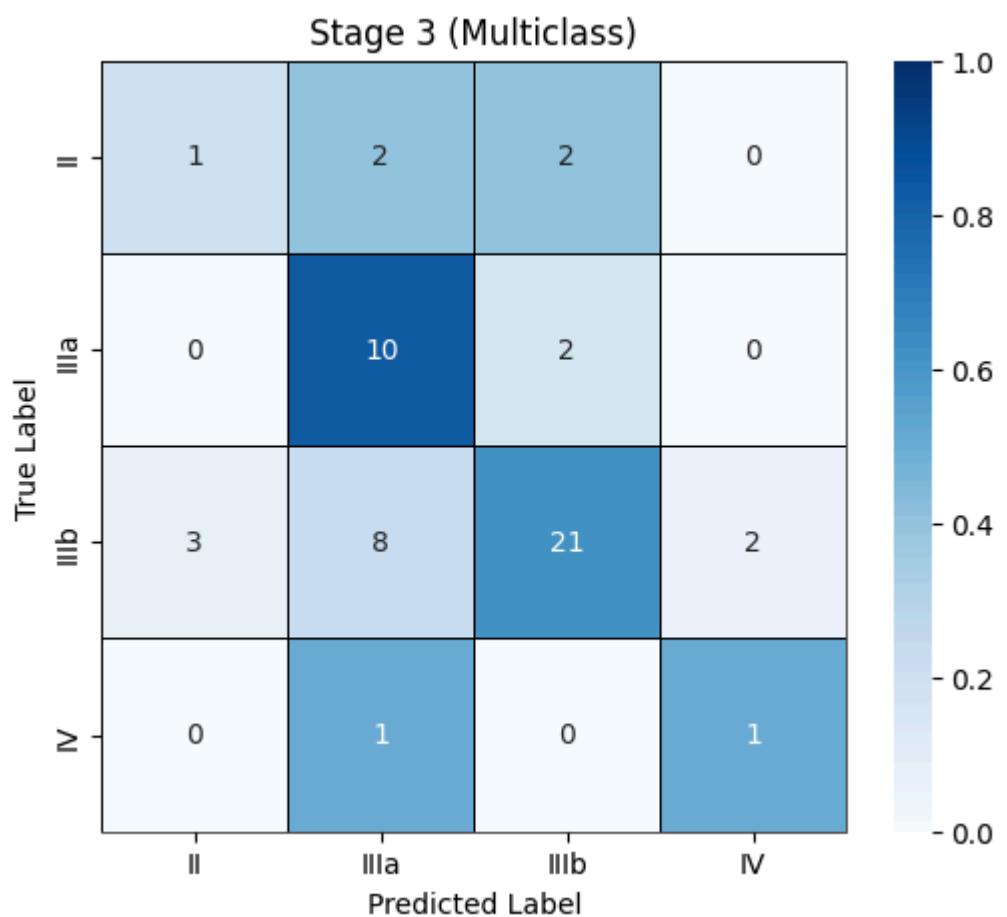
-- Stage 2 (DC vs Rest) --

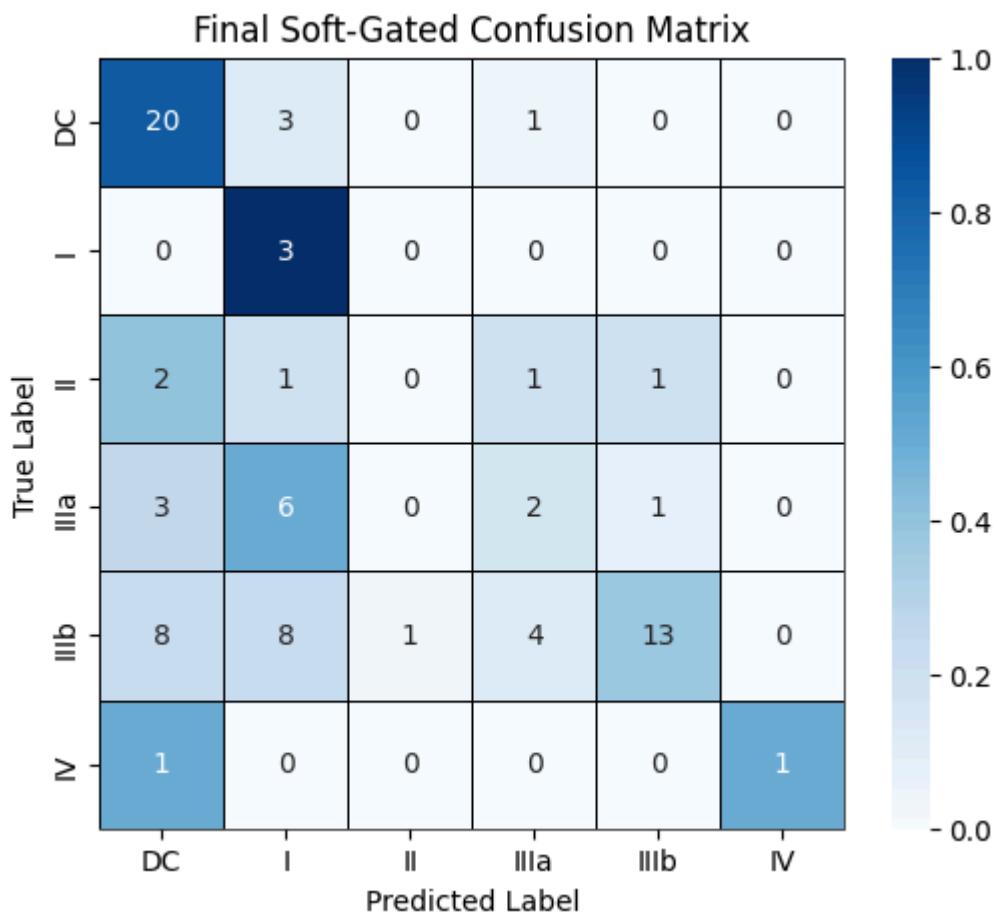
[Threshold Optimization] Best balanced_accuracy: 0.8146 at threshold=0.101

Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --





===== Fold 4 =====

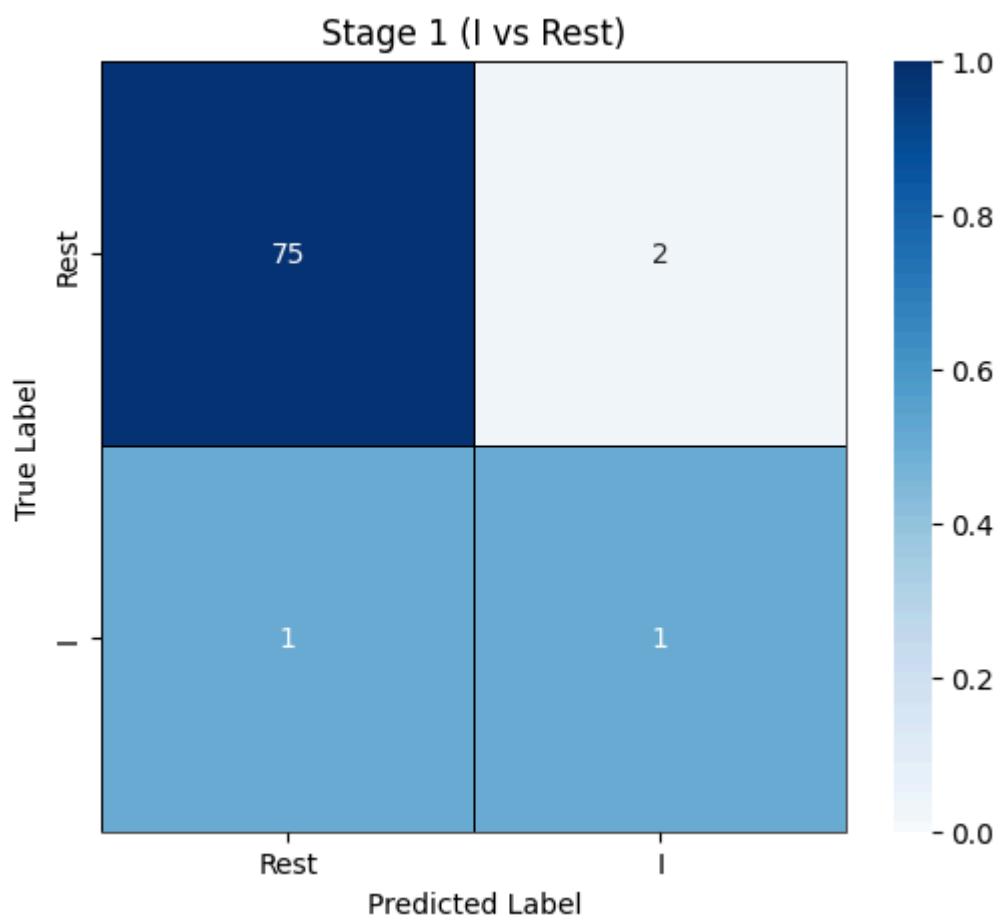
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

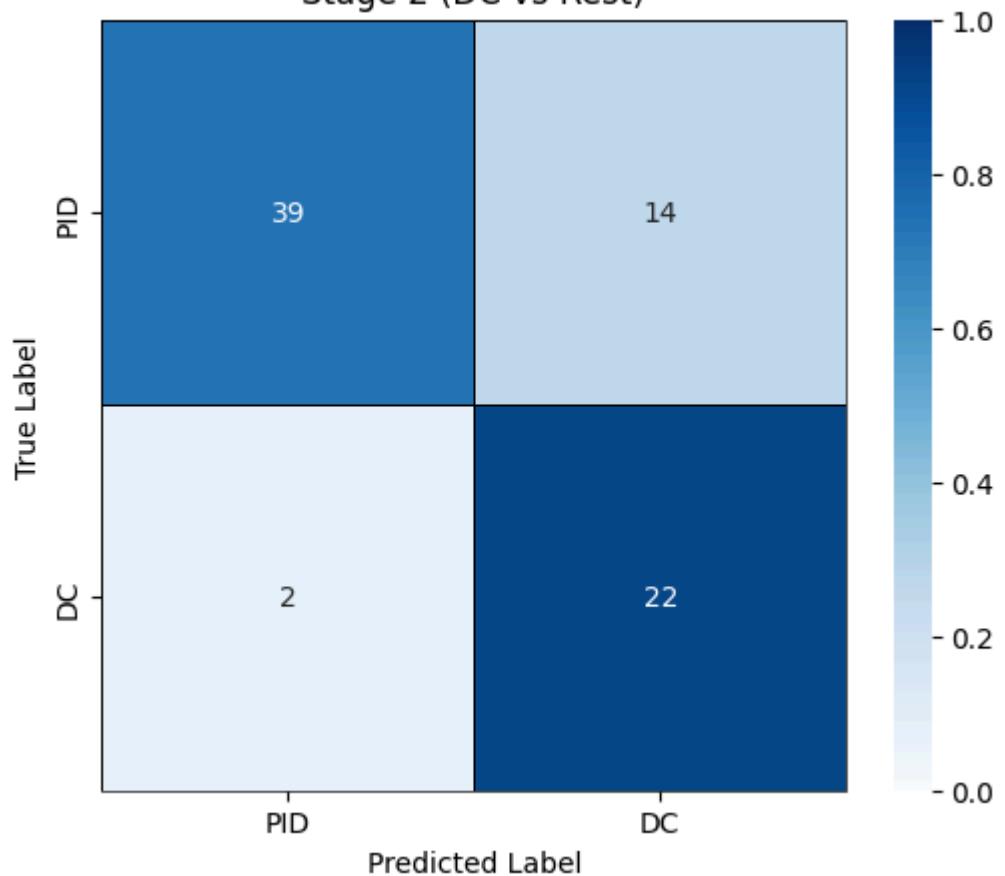
[Threshold Optimization] Best balanced_accuracy: 0.9872 at threshold=0.442



-- Stage 2 (DC vs Rest) --

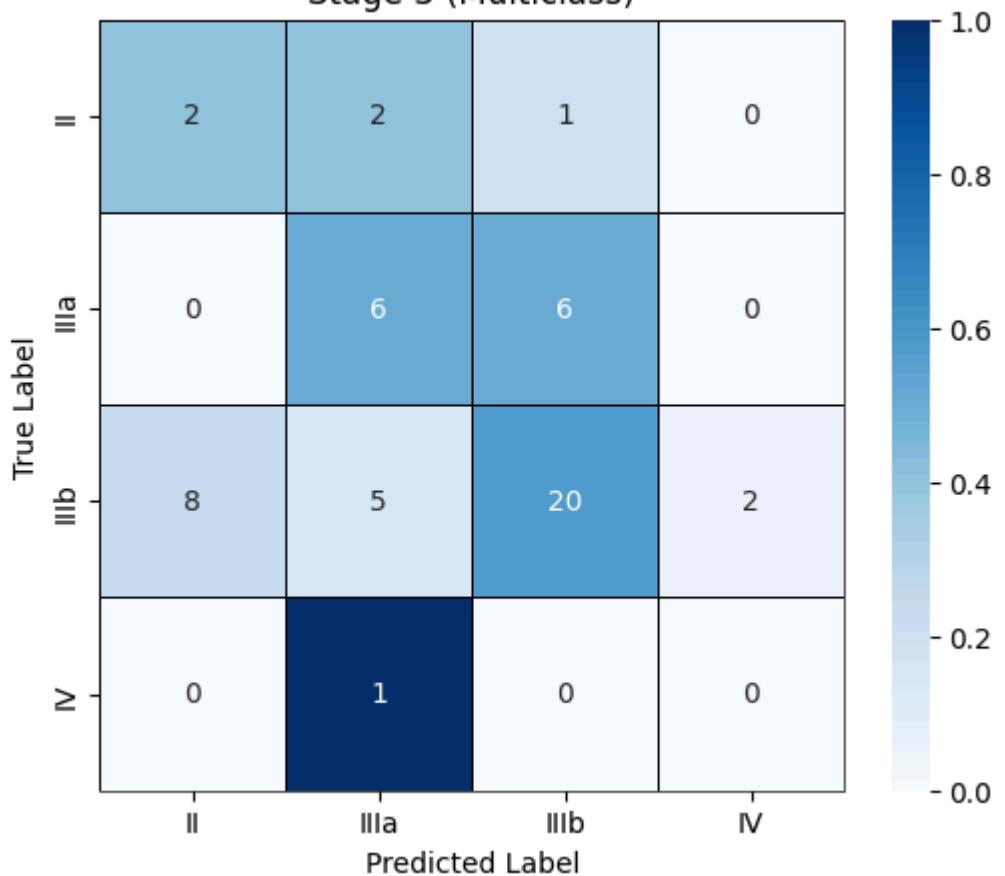
[Threshold Optimization] Best balanced_accuracy: 0.8273 at threshold=0.080

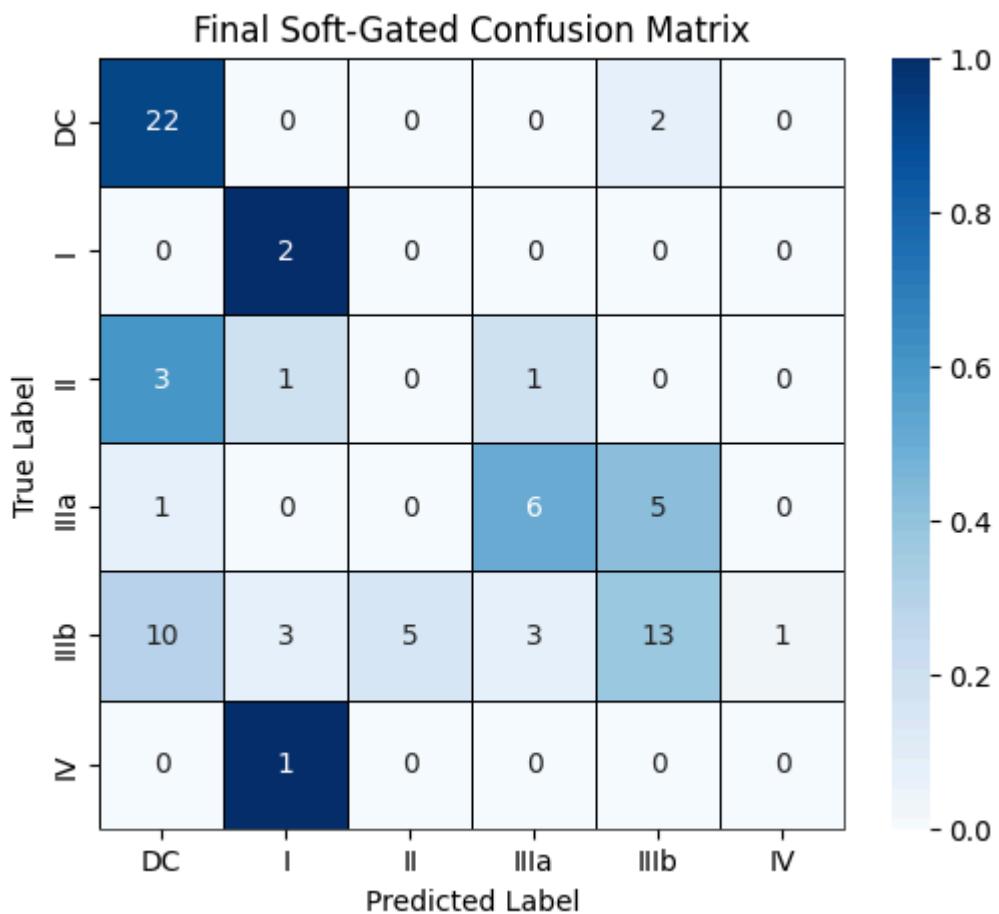
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 5 =====

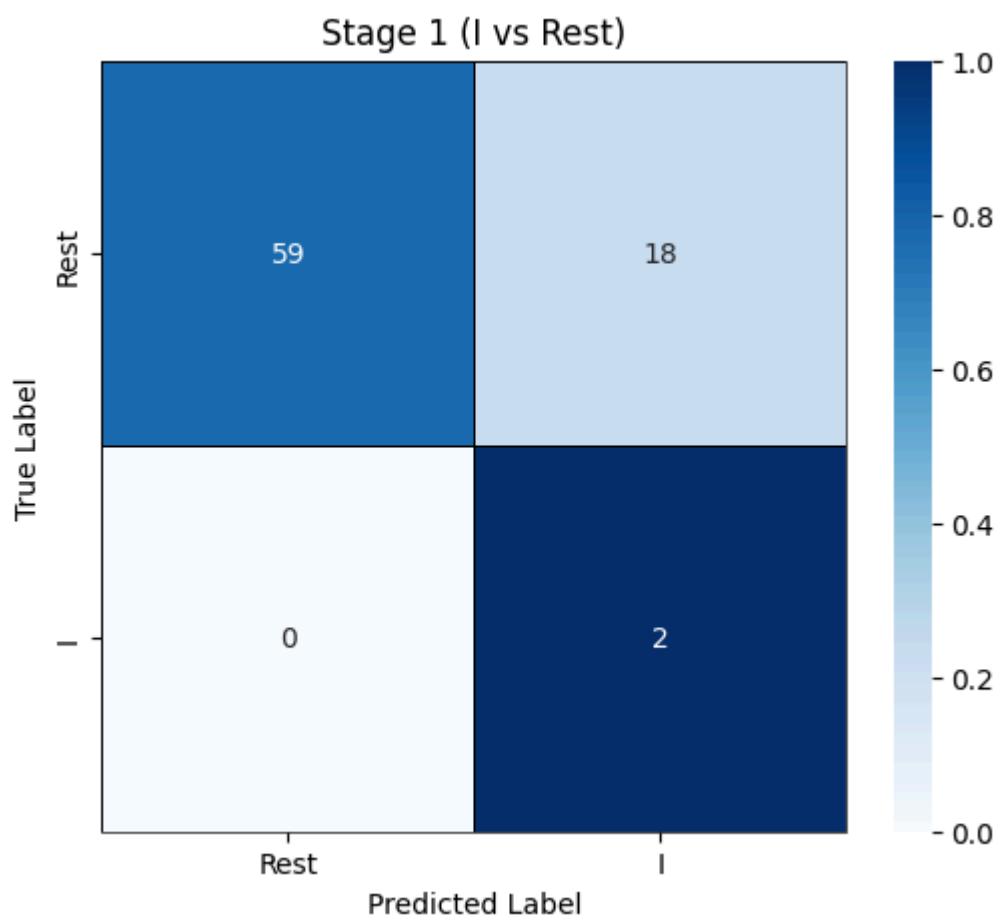
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

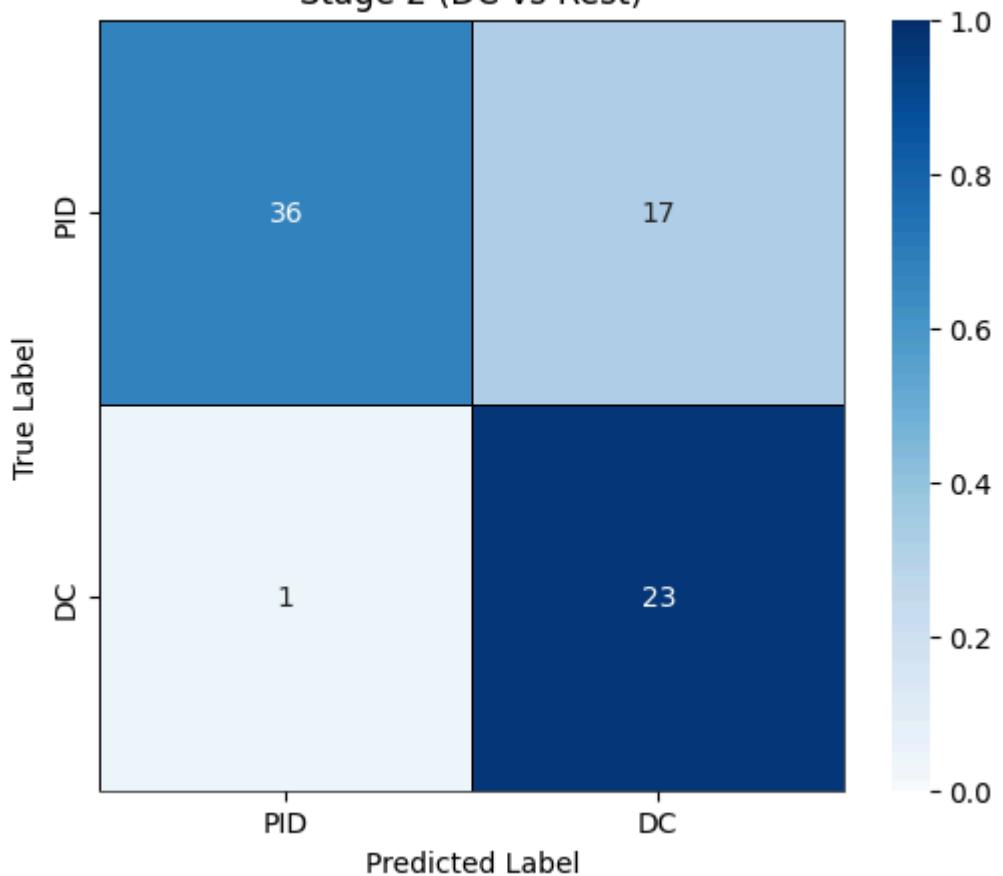
[Threshold Optimization] Best balanced_accuracy: 0.8526 at threshold=0.286



-- Stage 2 (DC vs Rest) --

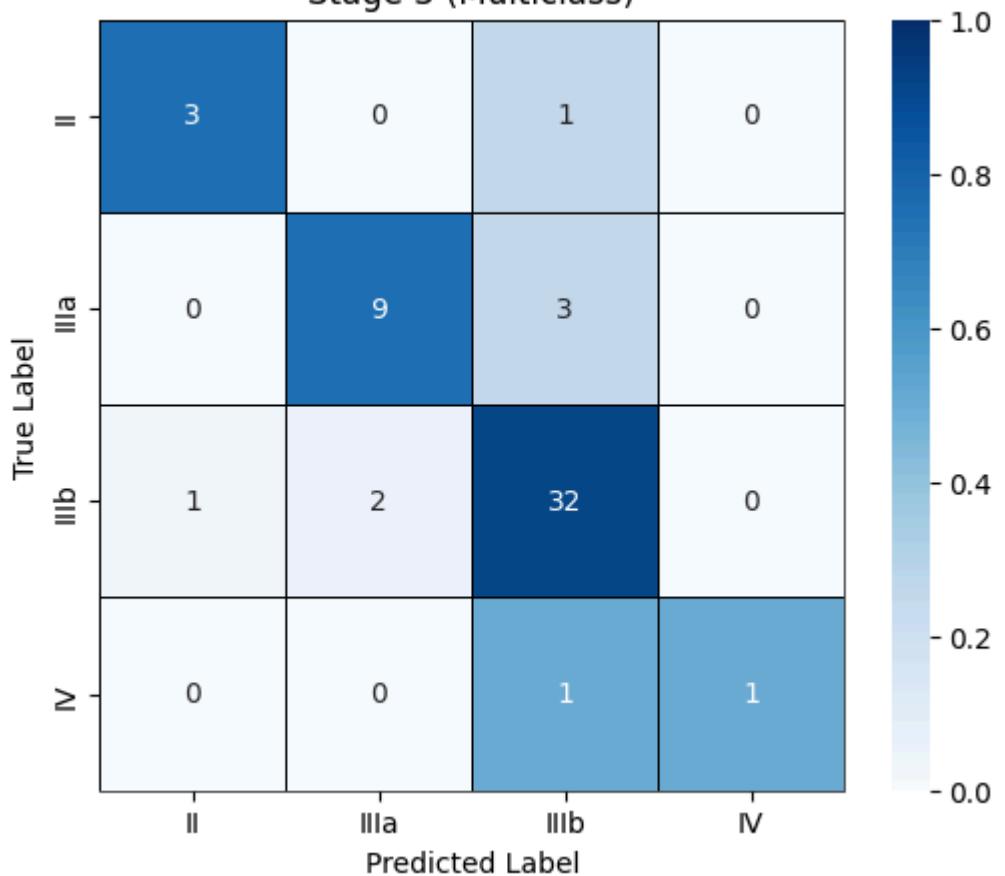
[Threshold Optimization] Best balanced_accuracy: 0.8747 at threshold=0.116

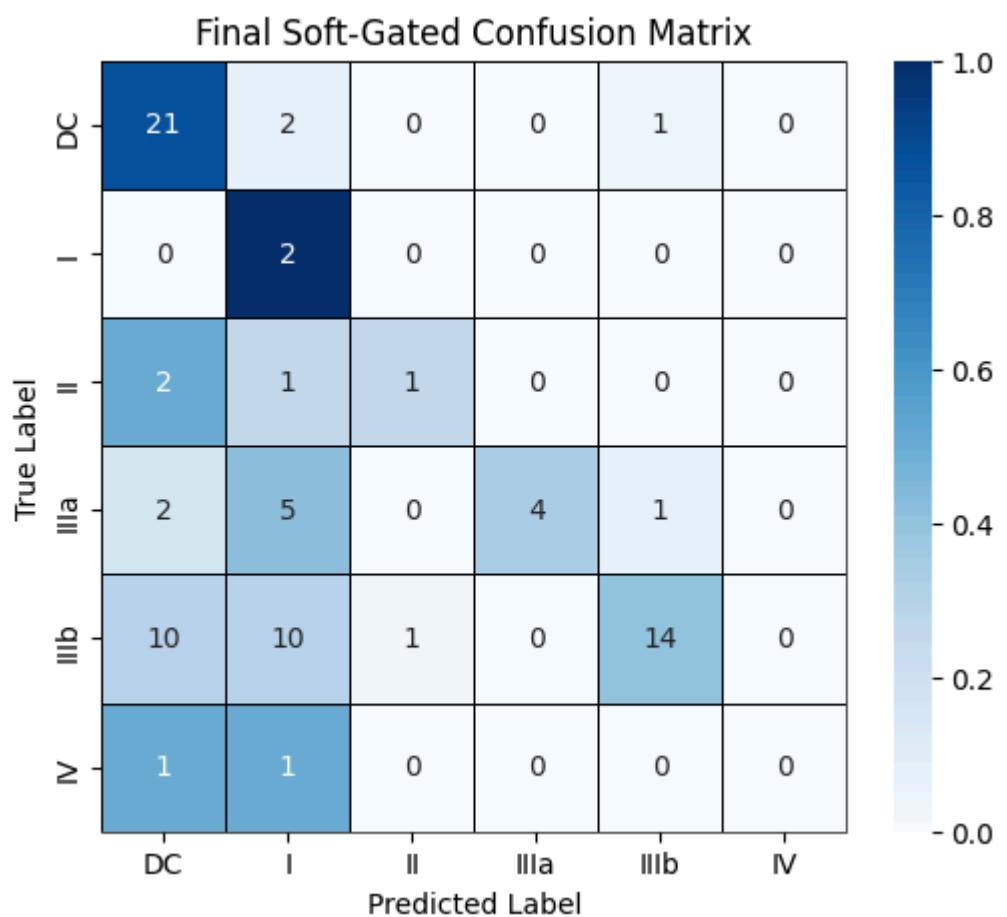
Stage 2 (DC vs Rest)



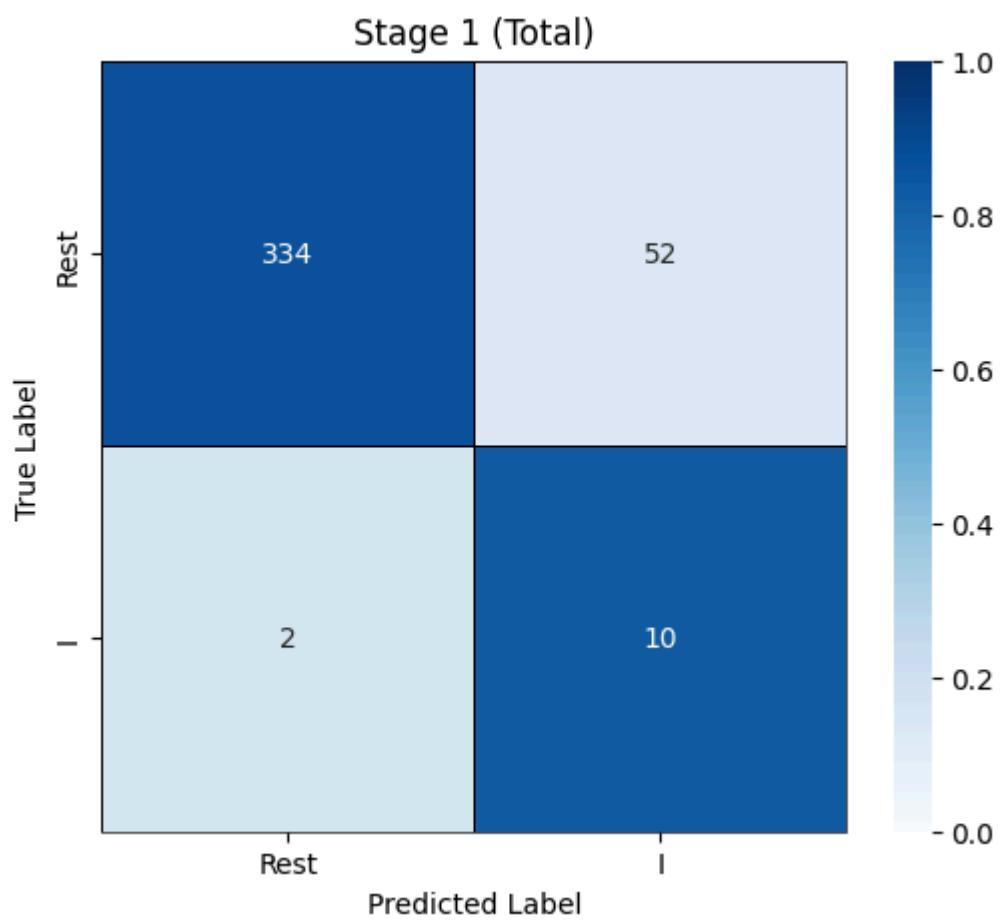
-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)

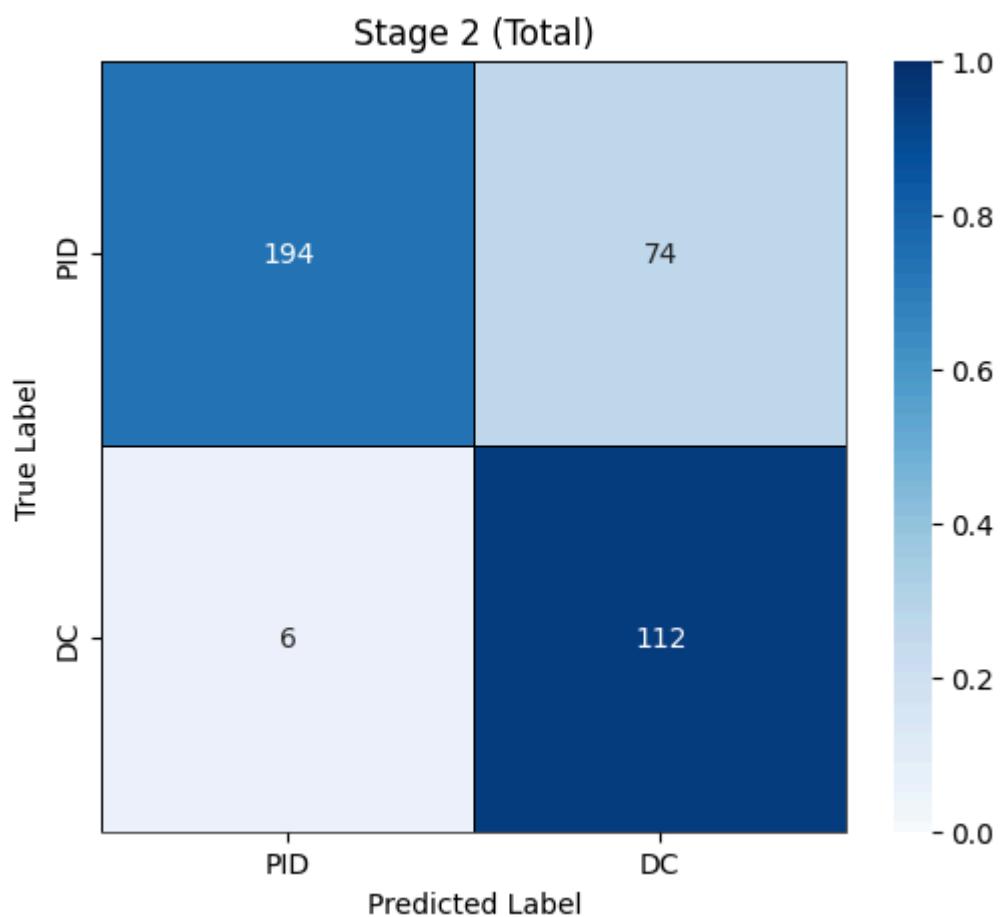




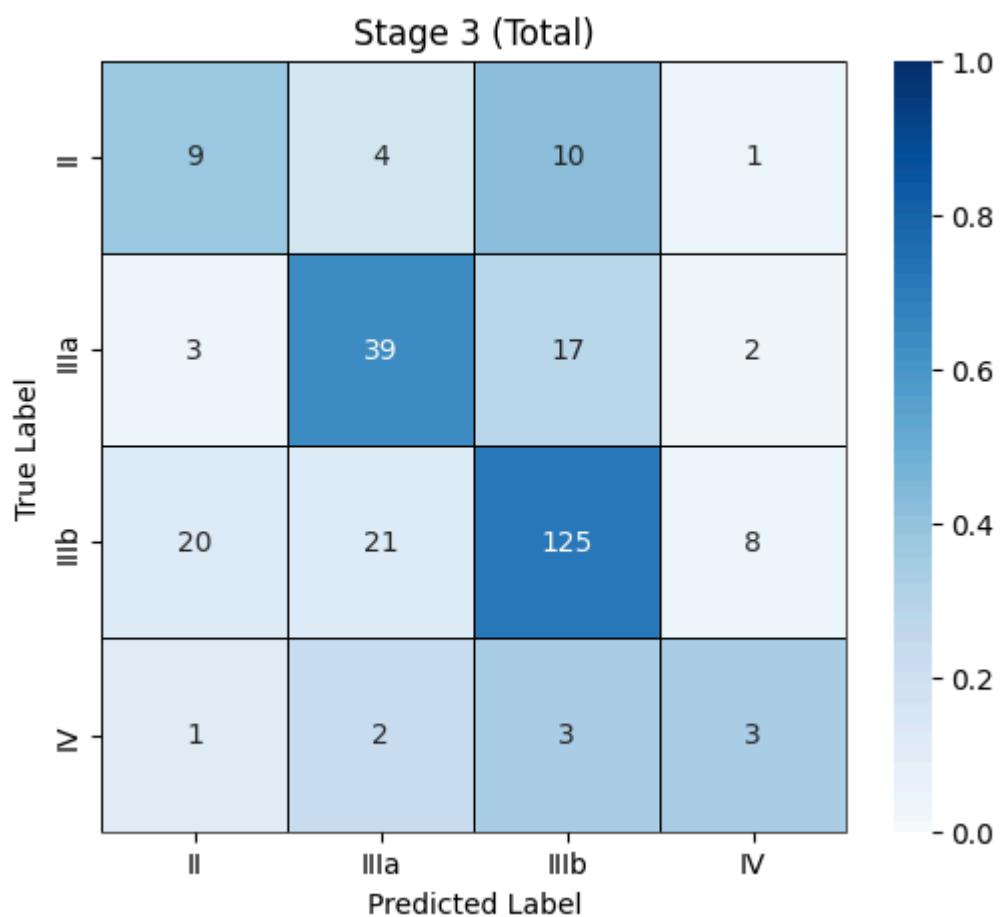
===== AGGREGATED CONFUSION MATRICES =====



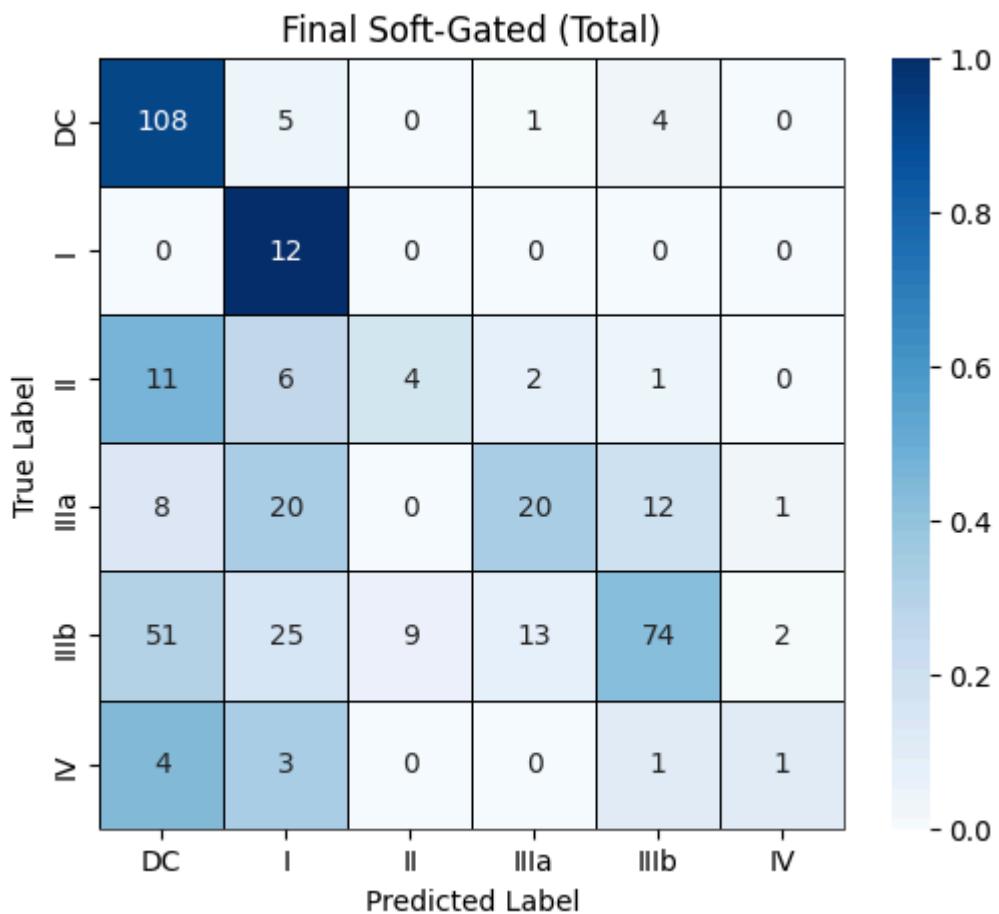
	precision	recall	f1-score	support
0	0.994	0.865	0.925	386
1	0.161	0.833	0.270	12
accuracy		0.864	0.864	398
macro avg	0.578	0.849	0.598	398
weighted avg	0.969	0.864	0.905	398



	precision	recall	f1-score	support
0	0.970	0.724	0.829	268
1	0.602	0.949	0.737	118
accuracy		0.793	0.793	386
macro avg	0.786	0.837	0.783	386
weighted avg	0.858	0.793	0.801	386



	precision	recall	f1-score	support
0	0.273	0.375	0.316	24
1	0.591	0.639	0.614	61
2	0.806	0.718	0.760	174
3	0.214	0.333	0.261	9
accuracy		0.657	0.657	268
macro avg	0.471	0.517	0.488	268
weighted avg	0.690	0.657	0.670	268



===== CV SUMMARY =====

precision recall f1-score support

DC	0.5934	0.9153	0.7200	118
I	0.1690	1.0000	0.2892	12
II	0.3077	0.1667	0.2162	24
IIIa	0.5556	0.3279	0.4124	61
IIIb	0.8043	0.4253	0.5564	174
IV	0.2500	0.1111	0.1538	9

accuracy		0.5503	398	
macro avg	0.4467	0.4910	0.3913	398
weighted avg	0.6420	0.5503	0.5452	398

Balanced Accuracy (Final Soft-Gated, aggregated): 0.4910313706411282

===== Accuracies =====

Per-fold balanced accuracy:

Stage 1: [0.7051282051282051, 0.9415584415584415, 0.8961038961038961, 0.737012987012987, 0.8831168831168832]

Stage 2: [0.8237154150197629, 0.8796296296296297, 0.8356918238993711, 0.8262578616352201, 0.8187893081761006]

Stage 3: [0.4997252747252747, 0.4303571428571429, 0.5377450980392158, 0.3678571428571429, 0.7285714285714285]

Mean per-fold balanced accuracy: [np.float64(0.8325840825840825), np.float64(0.8368168076720168), np.float64(0.5128512174100409)]

Aggregated balanced accuracy:

Stage 1: 0.8493091537132988

Stage 2: 0.8365165696939034

Stage 3: 0.5165171000565292

Final Soft-Gated: 0.4910313706411282

===== CV RUN 4/5 =====

Starting 5-split CV (3 train / 1 val / 1 test) on 398 samples...

===== Fold 1 =====

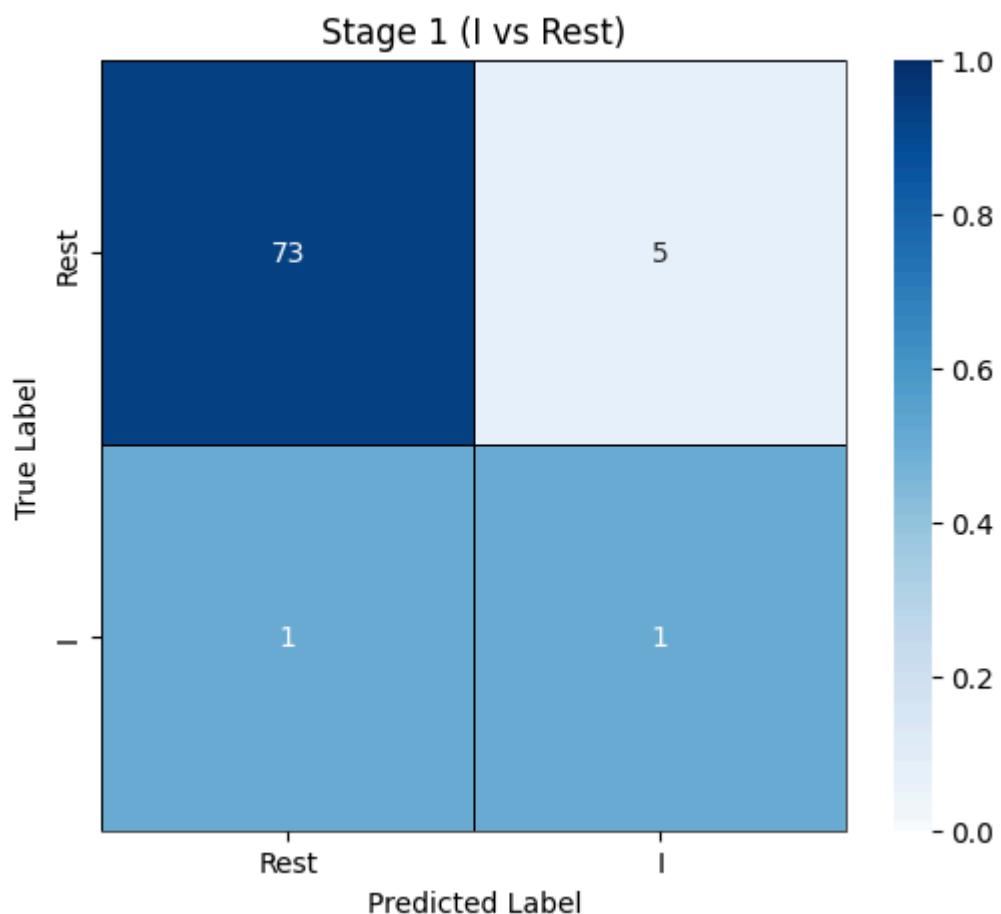
Train: 238, Val: 80, Test: 80

These features will be dropped:

['C11', 'C101', 'C64', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']

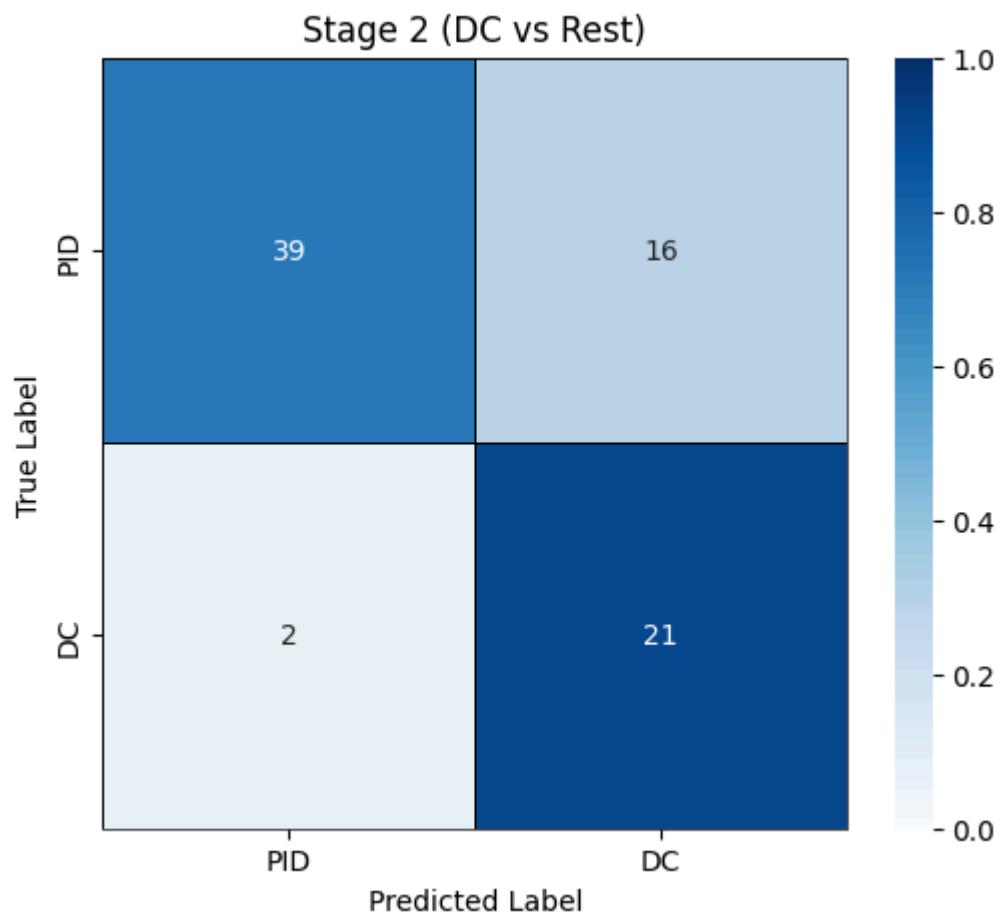
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 0.9610 at threshold=0.437



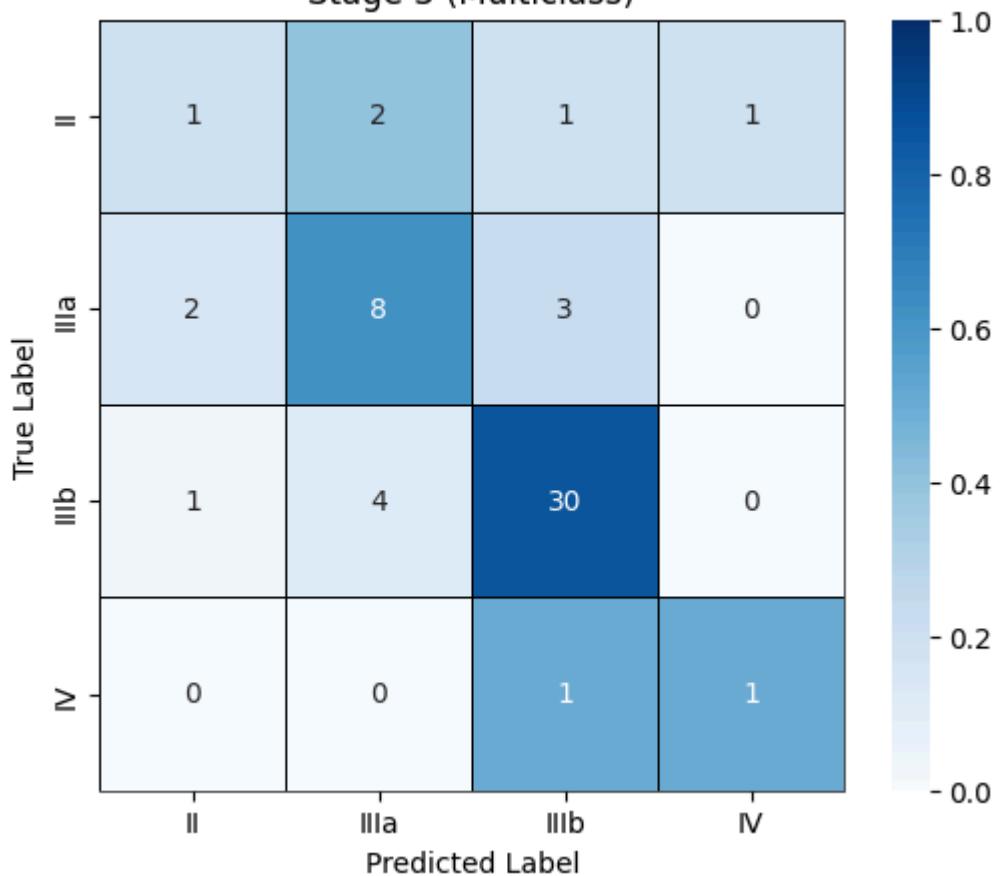
-- Stage 2 (DC vs Rest) --

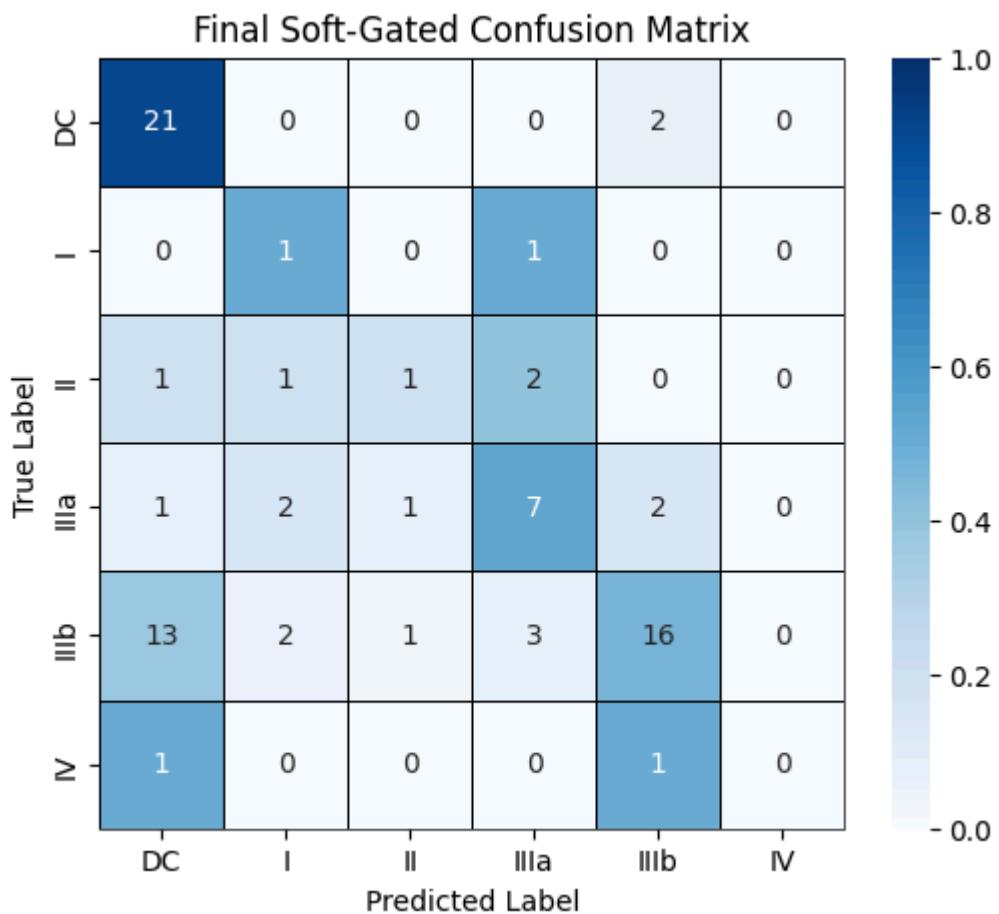
[Threshold Optimization] Best balanced_accuracy: 0.8575 at threshold=0.291



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 2 =====

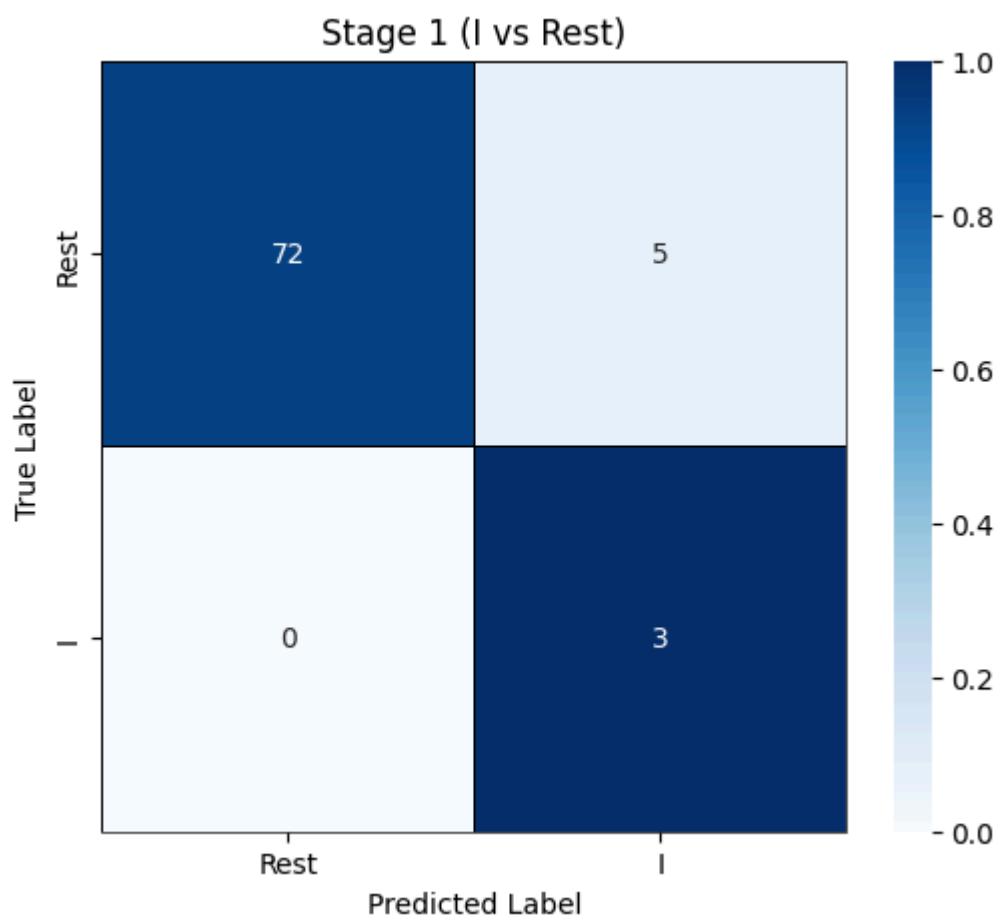
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

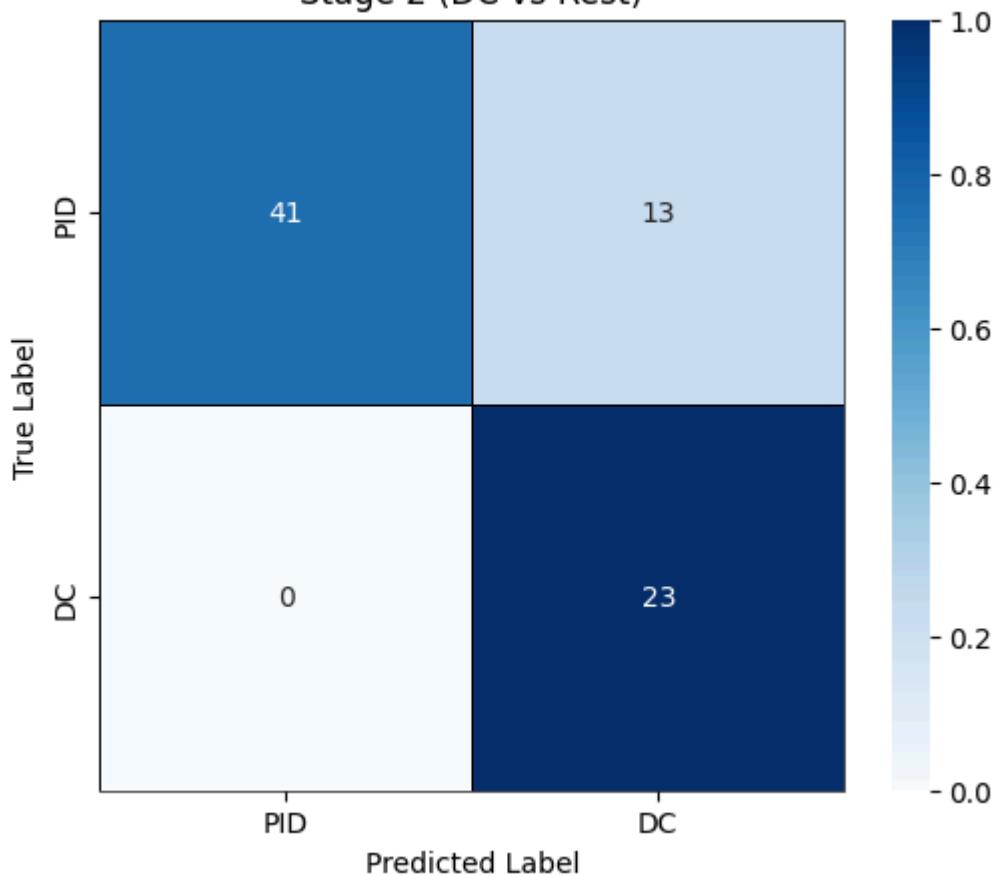
[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.608



-- Stage 2 (DC vs Rest) --

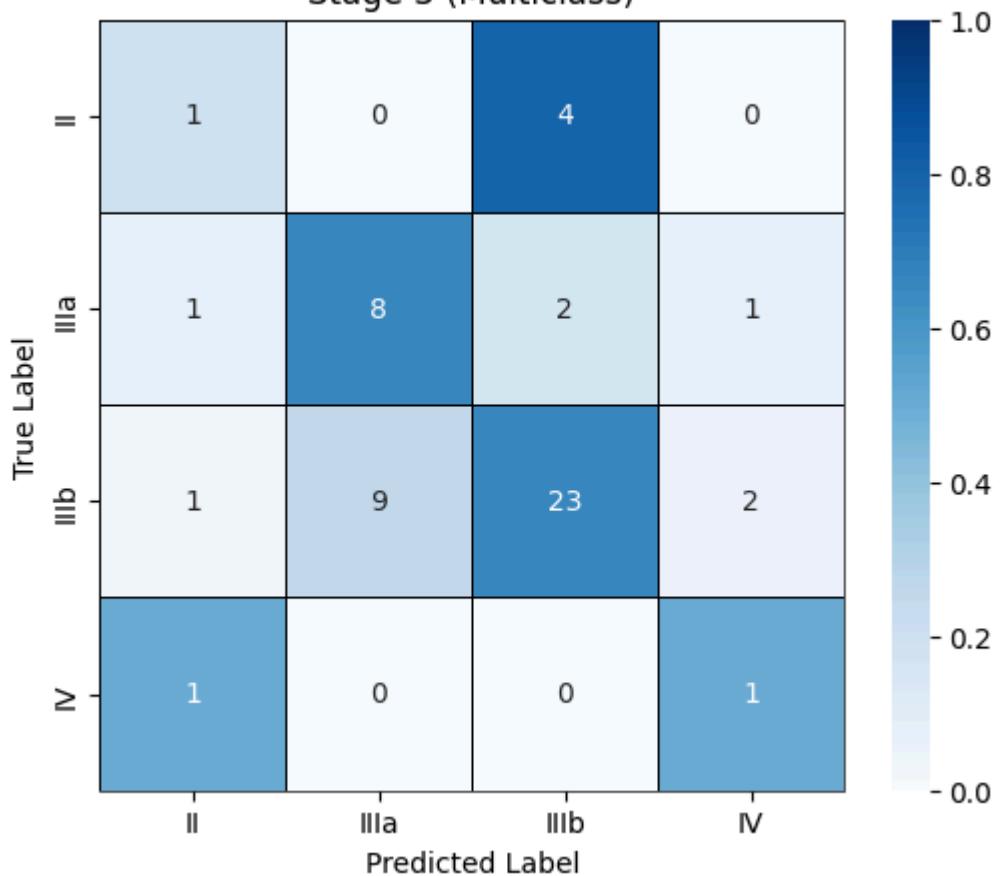
[Threshold Optimization] Best balanced_accuracy: 0.8565 at threshold=0.231

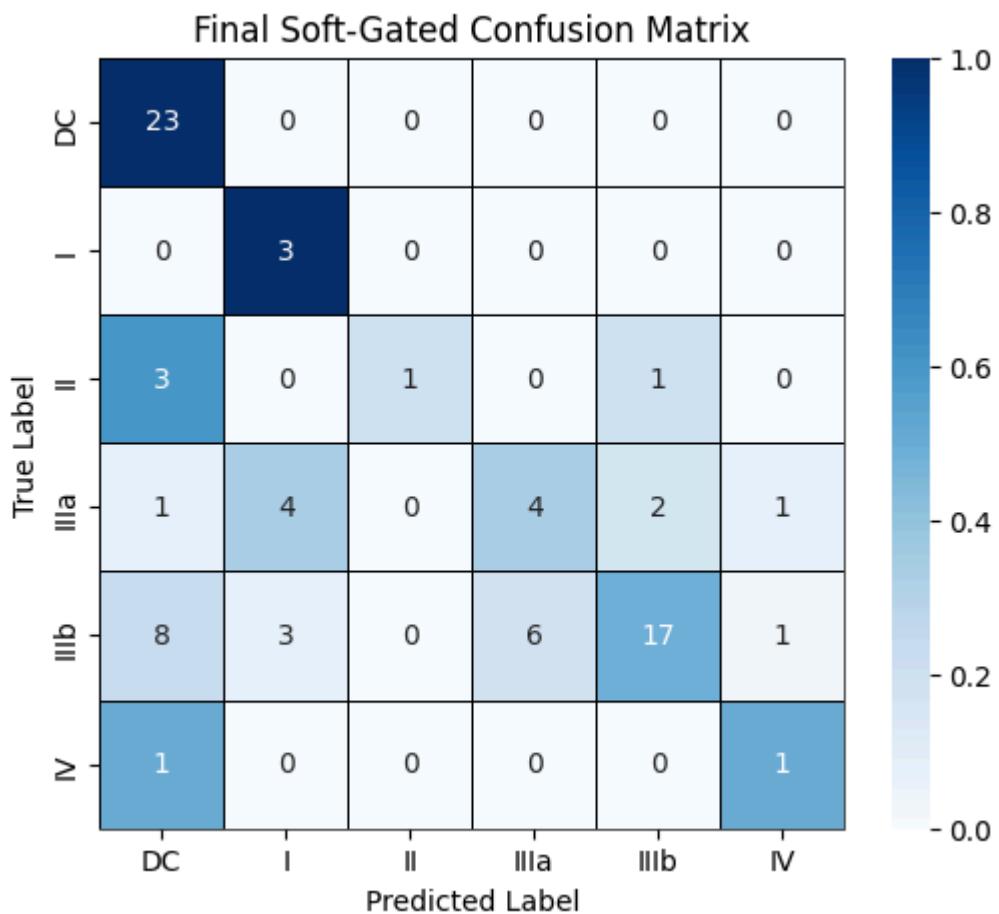
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 3 =====

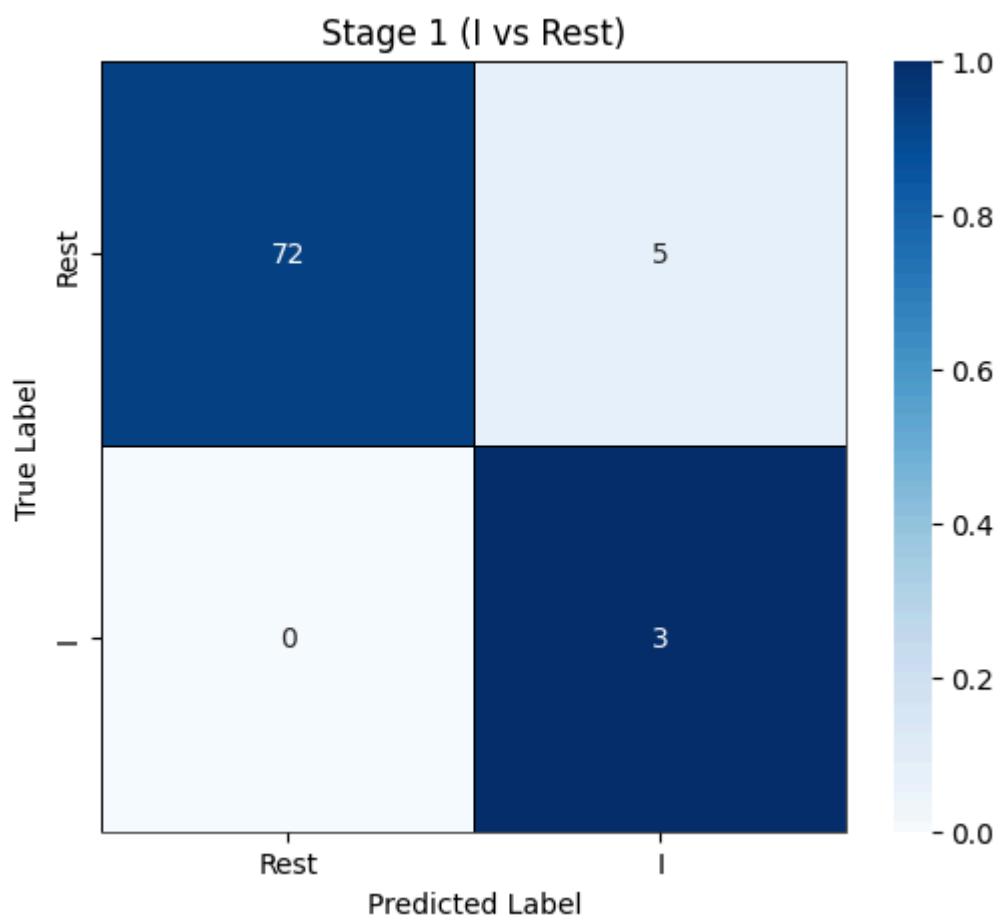
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'MC6. DP T cells', 'C64', 'C100', 'C126', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

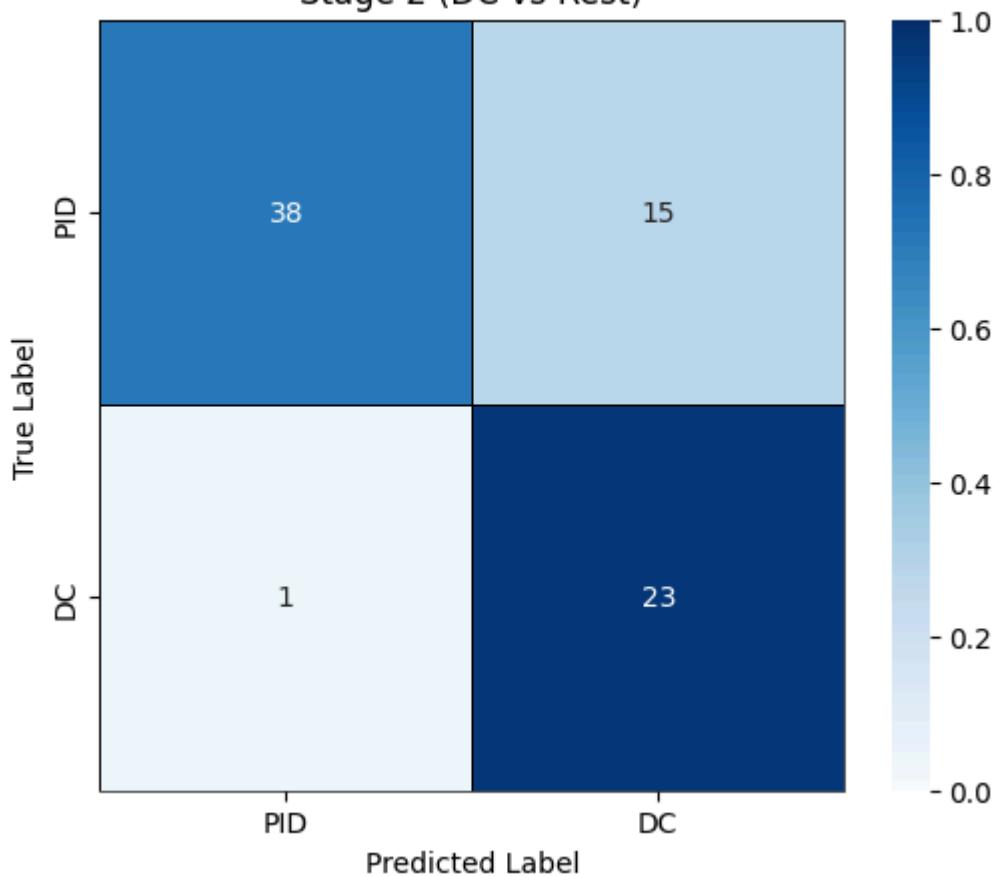
[Threshold Optimization] Best balanced_accuracy: 0.9679 at threshold=0.347



-- Stage 2 (DC vs Rest) --

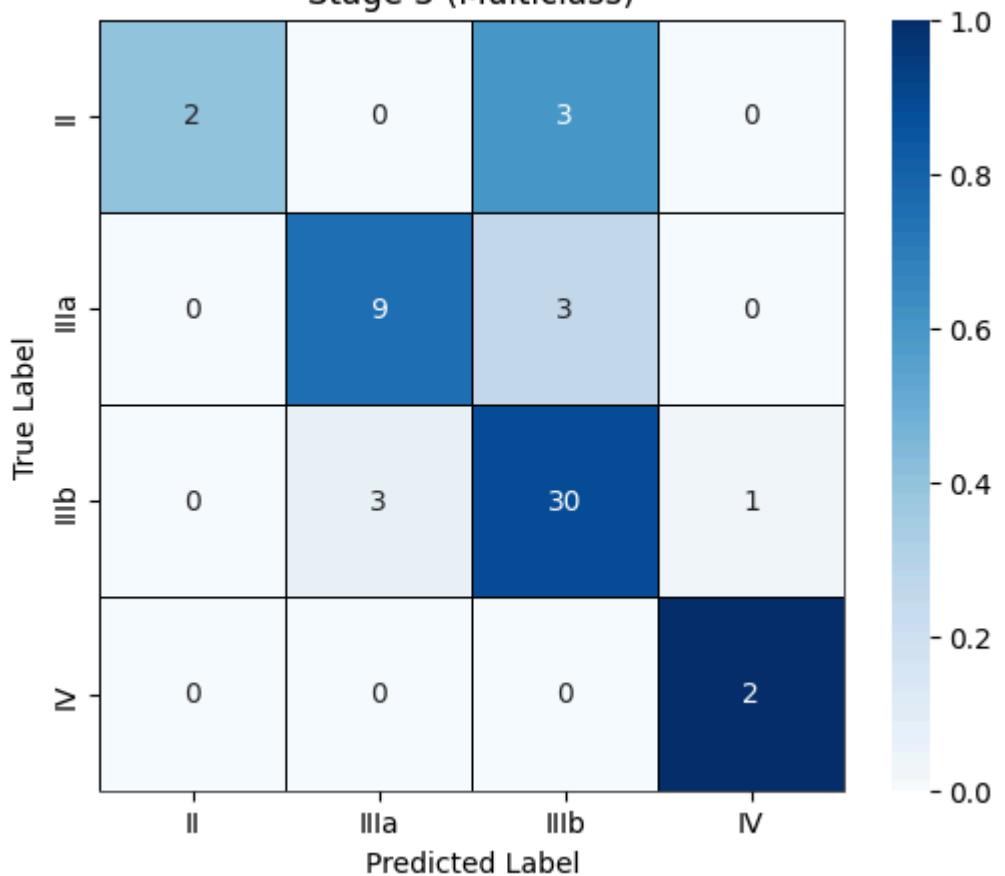
[Threshold Optimization] Best balanced_accuracy: 0.7964 at threshold=0.106

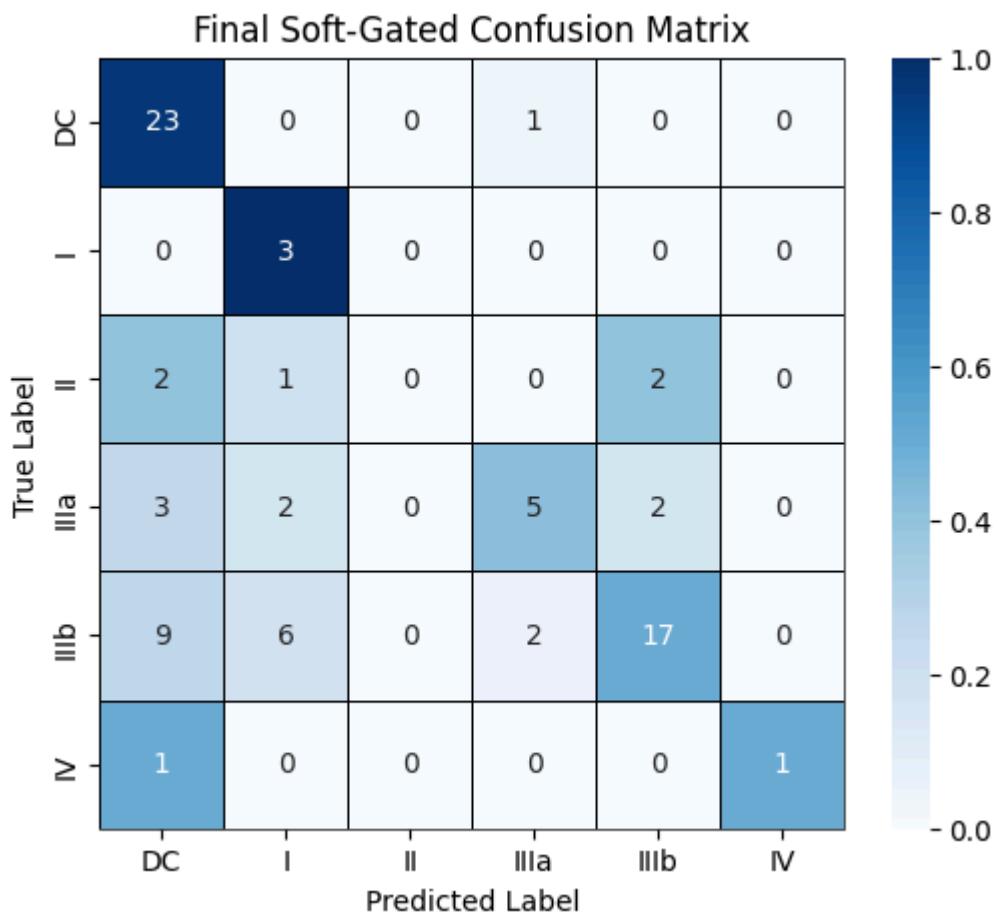
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 4 =====

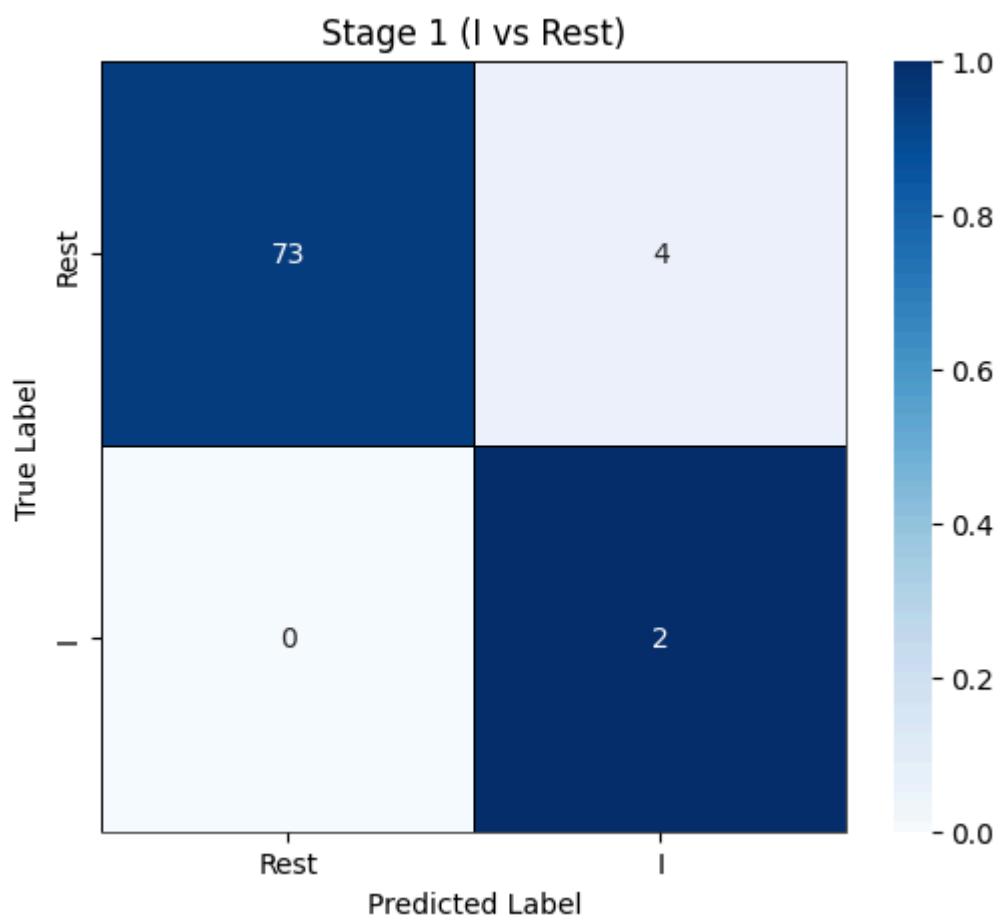
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

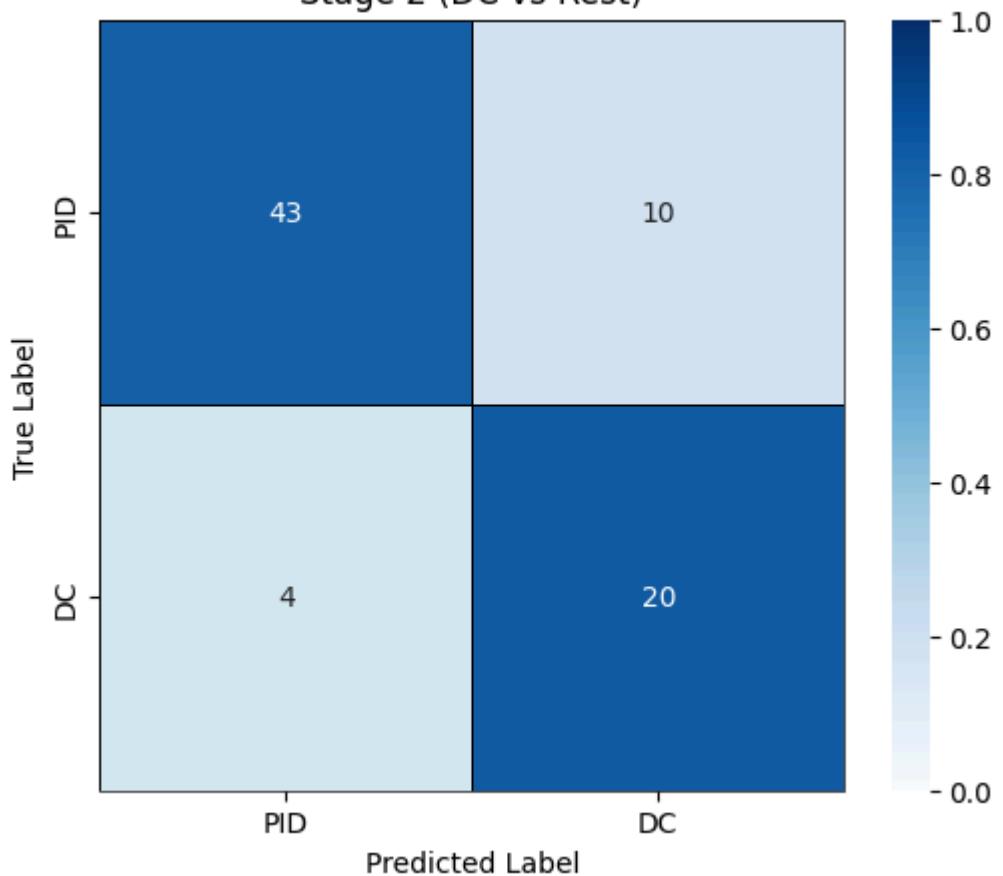
[Threshold Optimization] Best balanced_accuracy: 0.9551 at threshold=0.266



-- Stage 2 (DC vs Rest) --

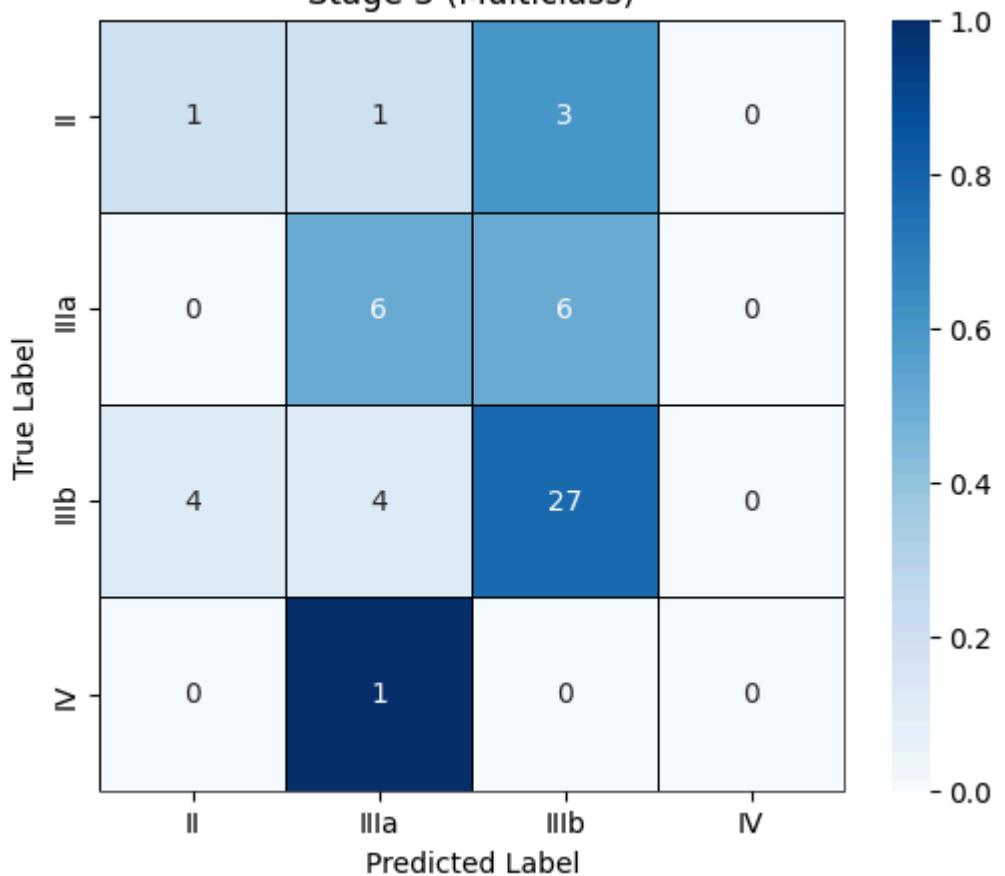
[Threshold Optimization] Best balanced_accuracy: 0.8474 at threshold=0.296

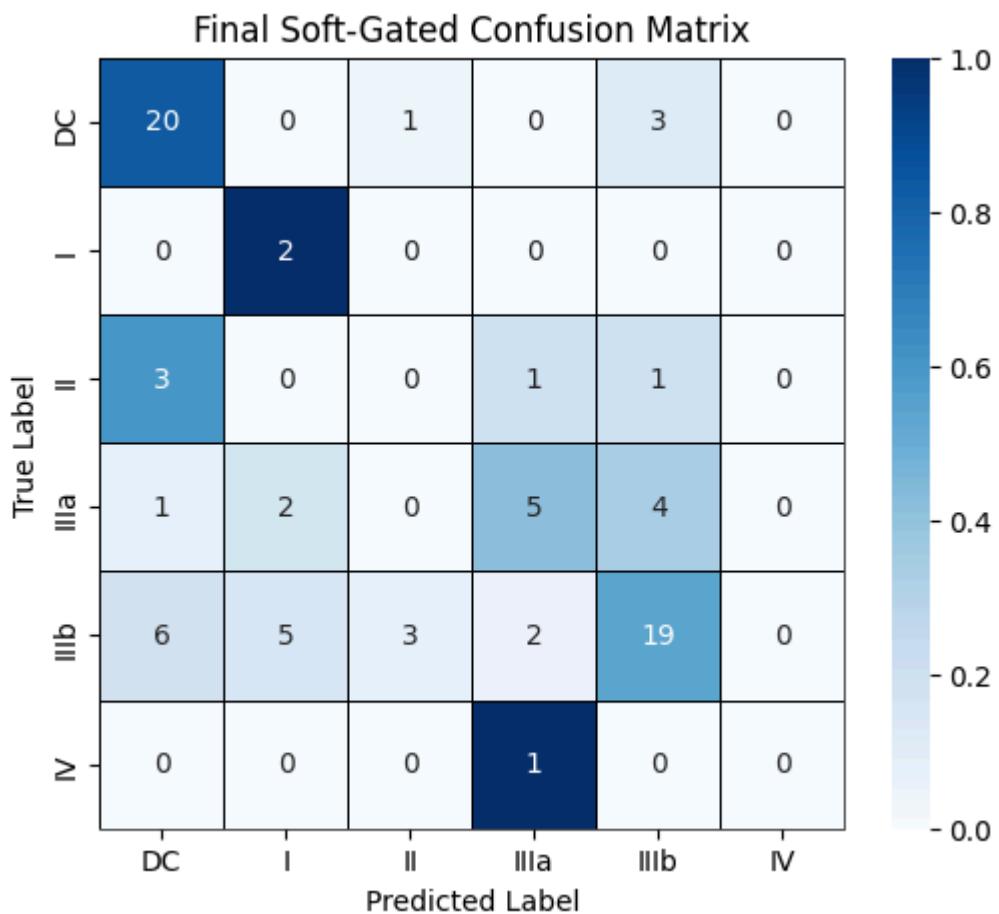
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 5 =====

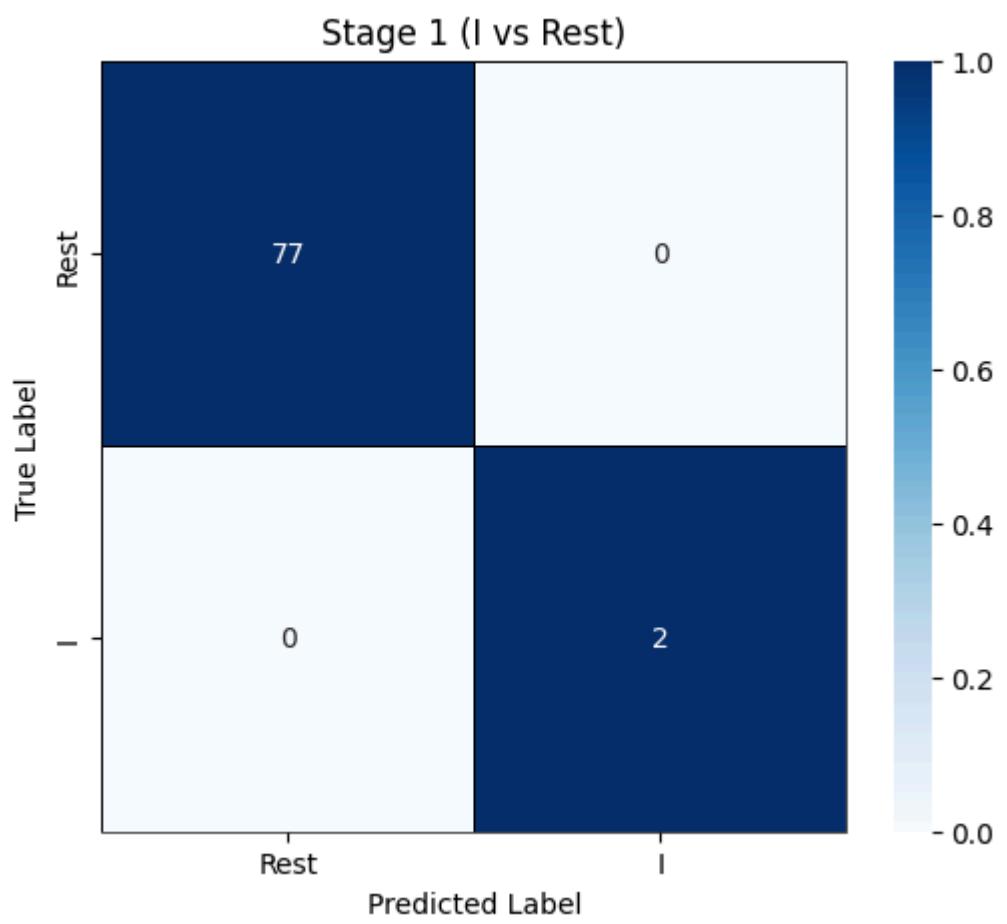
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']`

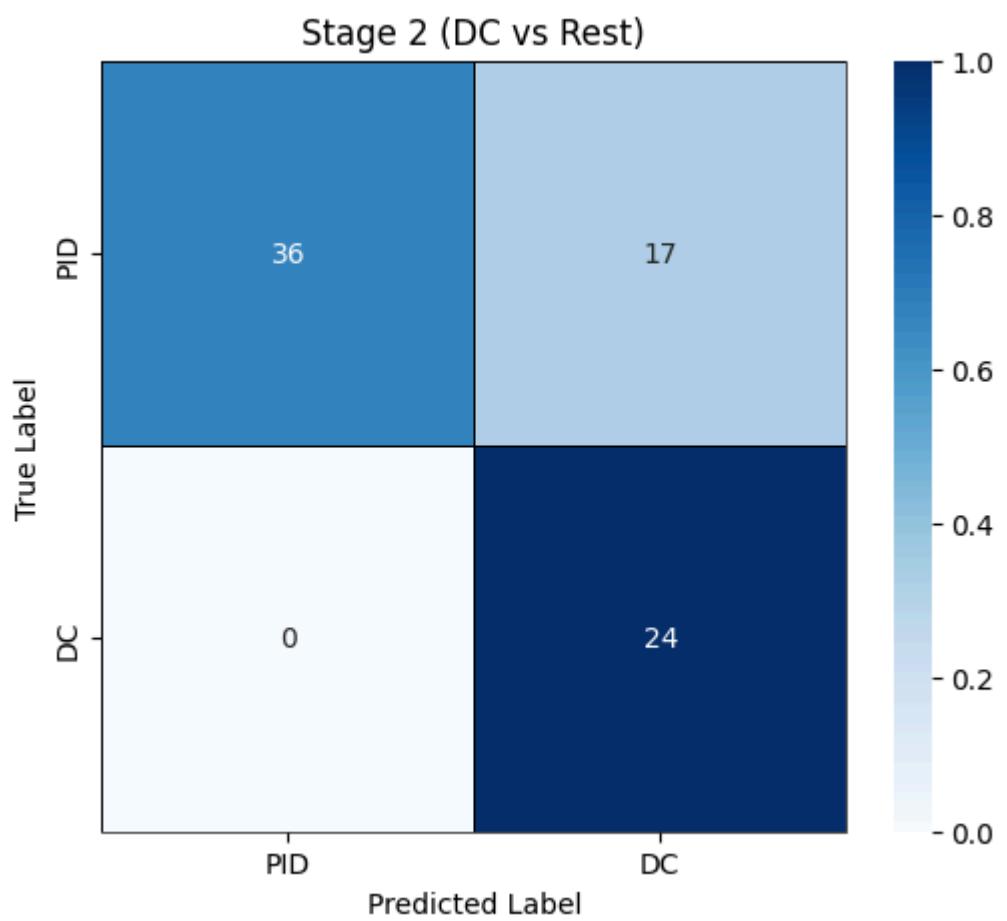
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.432



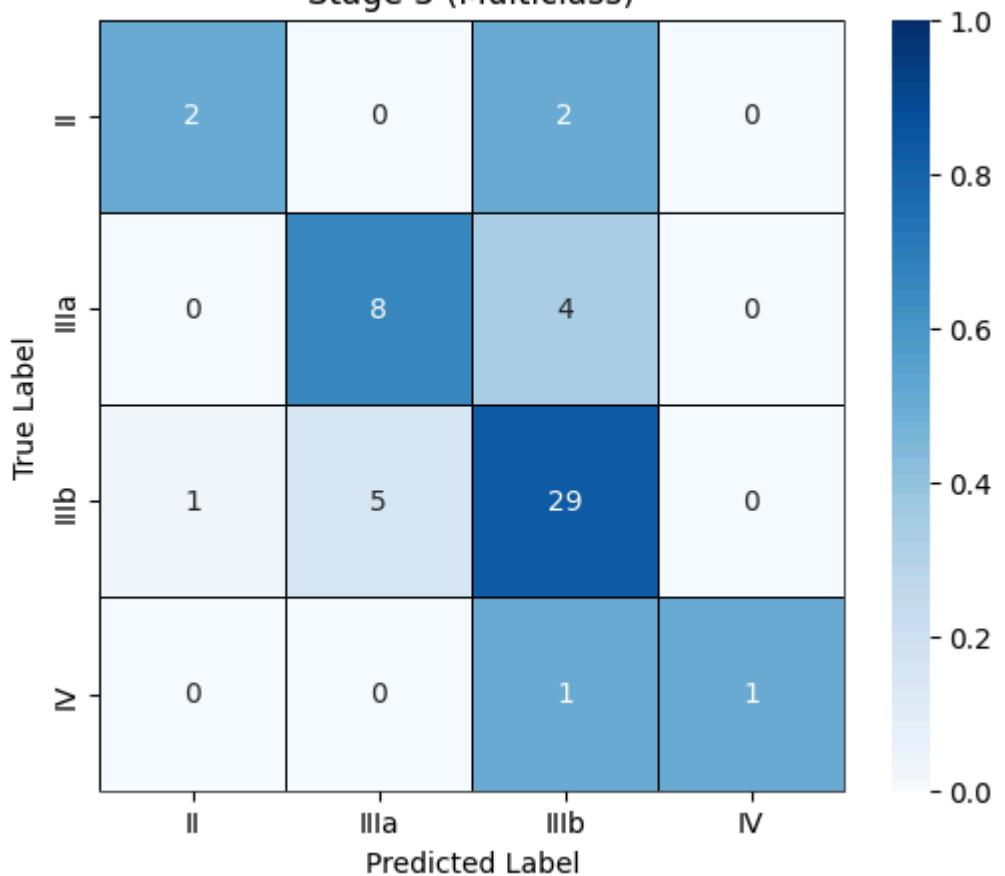
-- Stage 2 (DC vs Rest) --

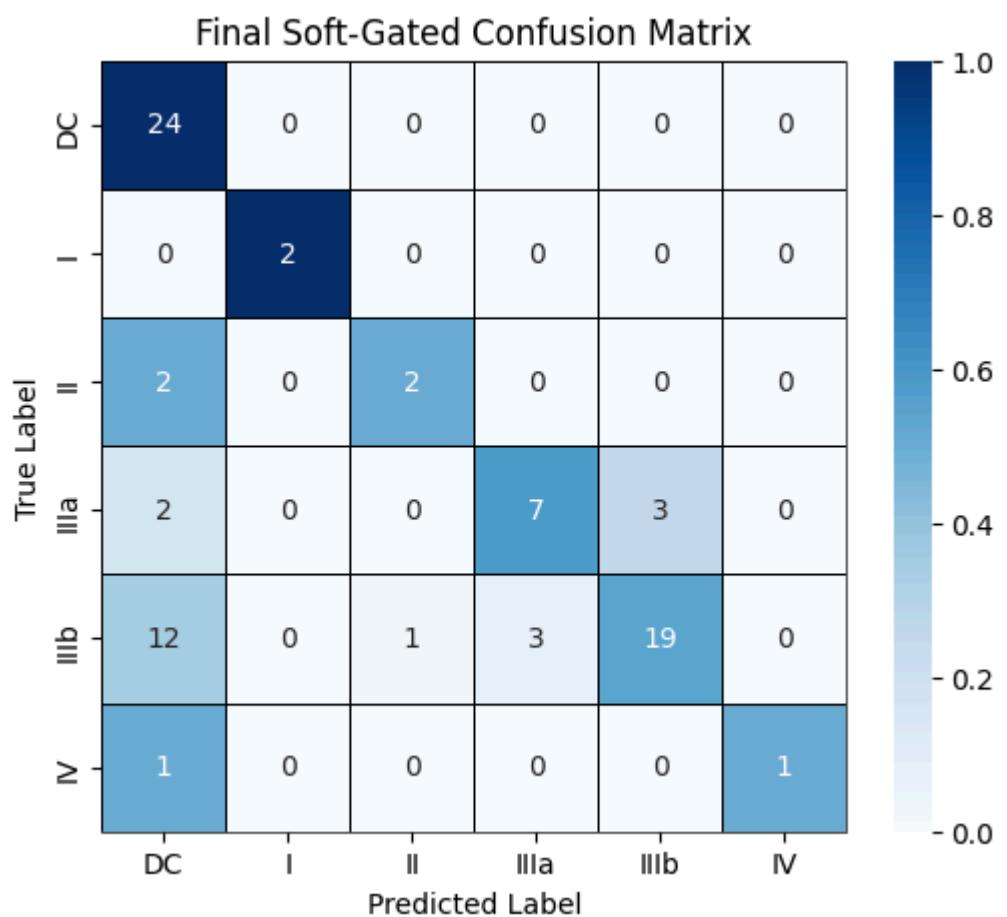
[Threshold Optimization] Best balanced_accuracy: 0.8656 at threshold=0.090



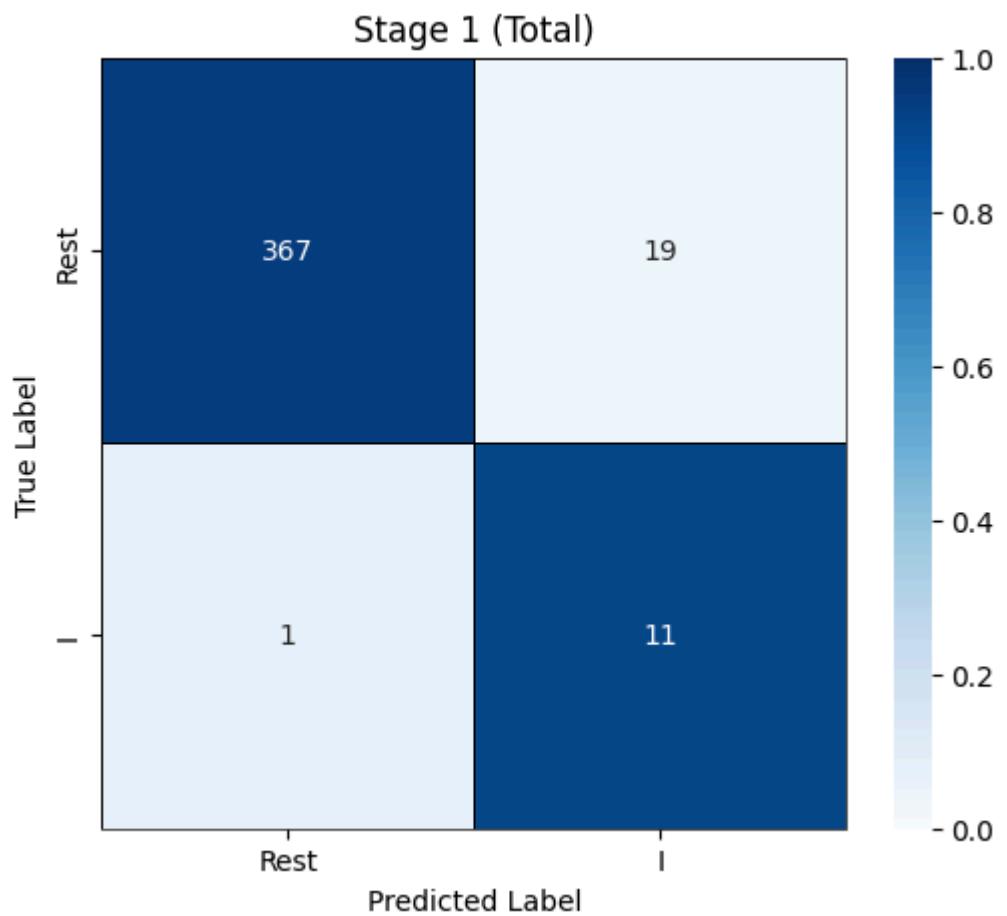
-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)

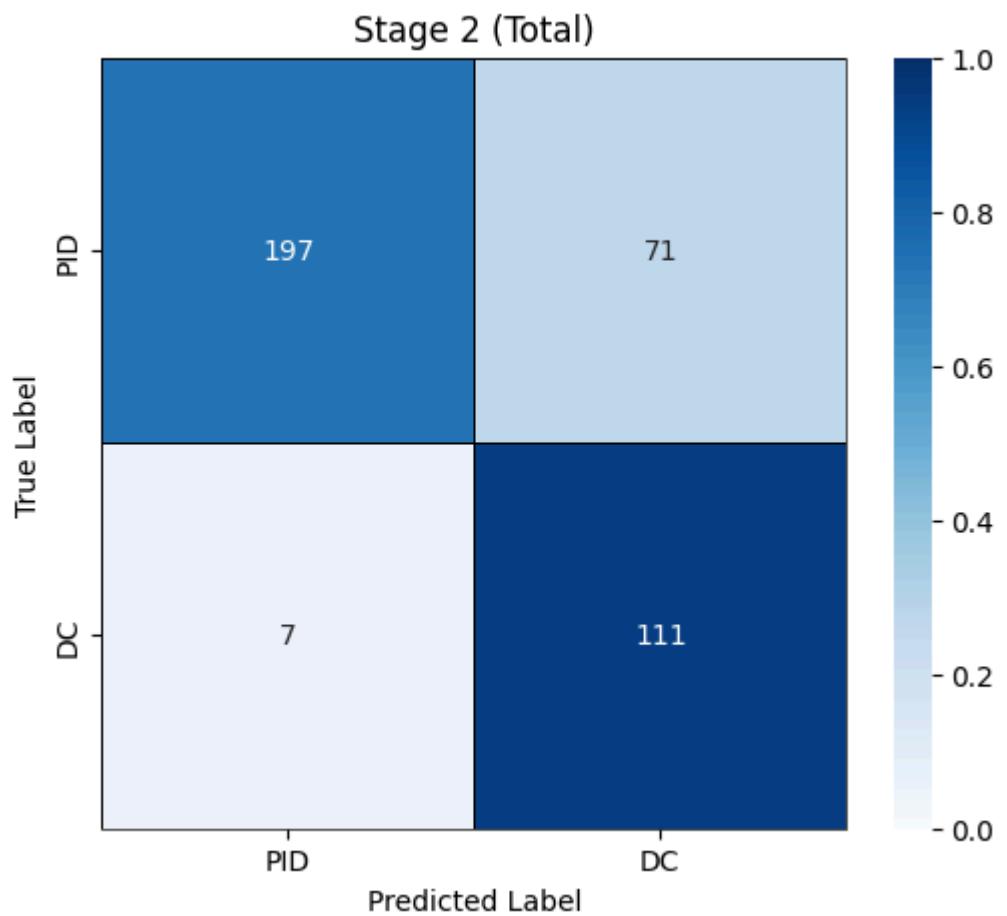




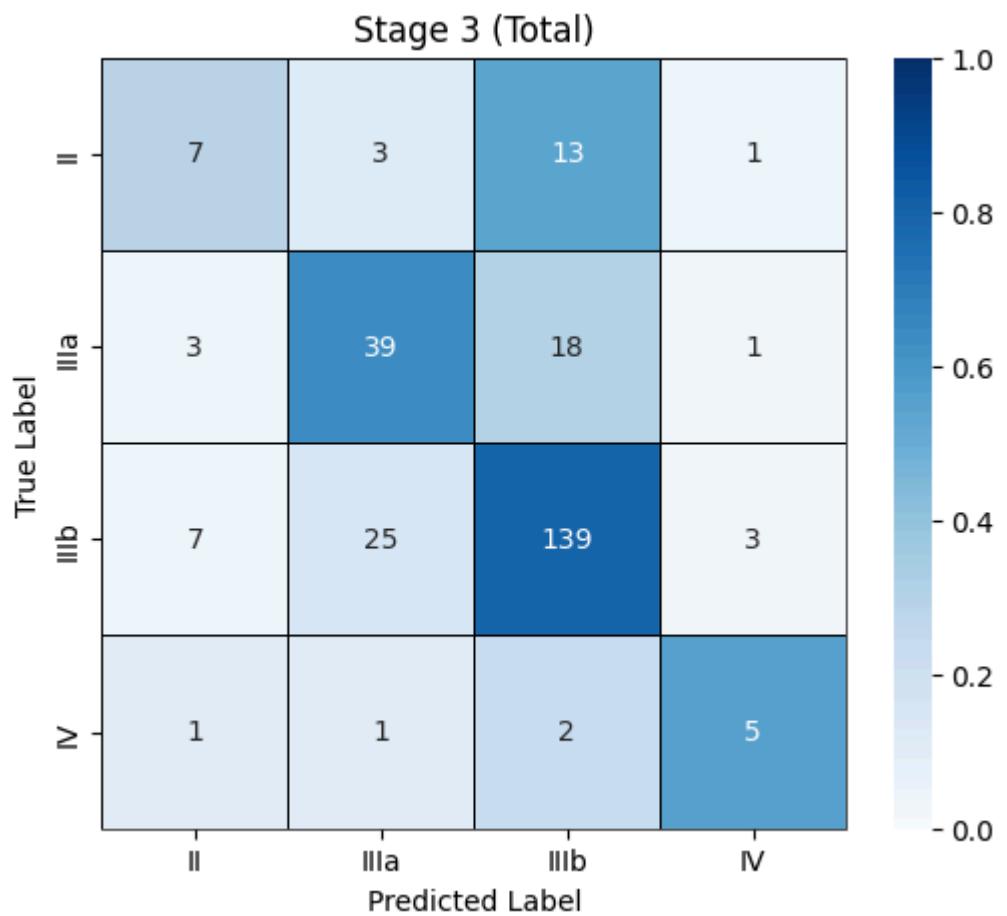
===== AGGREGATED CONFUSION MATRICES =====



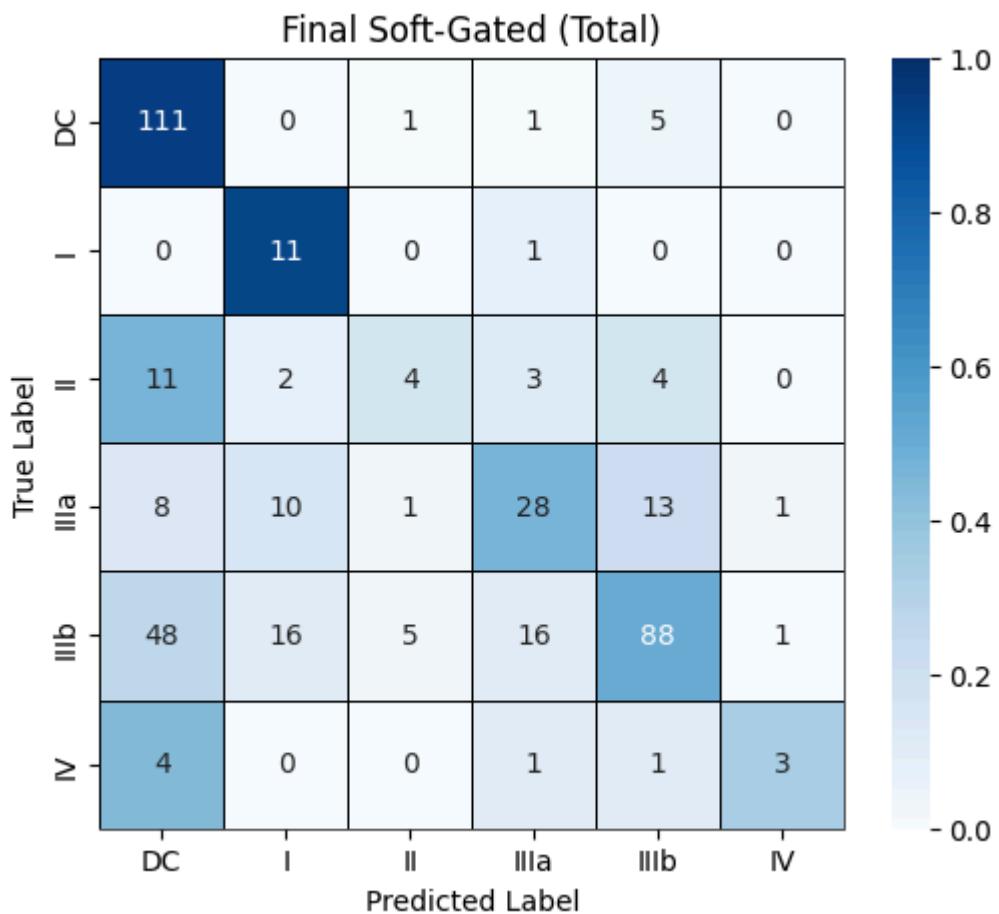
	precision	recall	f1-score	support
0	0.997	0.951	0.973	386
1	0.367	0.917	0.524	12
accuracy		0.950	0.950	398
macro avg	0.682	0.934	0.749	398
weighted avg	0.978	0.950	0.960	398



	precision	recall	f1-score	support
0	0.966	0.735	0.835	268
1	0.610	0.941	0.740	118
accuracy		0.798	0.798	386
macro avg	0.788	0.838	0.787	386
weighted avg	0.857	0.798	0.806	386



	precision	recall	f1-score	support
0	0.389	0.292	0.333	24
1	0.574	0.639	0.605	61
2	0.808	0.799	0.803	174
3	0.500	0.556	0.526	9
accuracy		0.709	0.709	268
macro avg	0.568	0.571	0.567	268
weighted avg	0.707	0.709	0.707	268



===== CV SUMMARY =====

precision recall f1-score support

DC	0.6099	0.9407	0.7400	118
I	0.2821	0.9167	0.4314	12
II	0.3636	0.1667	0.2286	24
IIIa	0.5600	0.4590	0.5045	61
IIIb	0.7928	0.5057	0.6175	174
IV	0.6000	0.3333	0.4286	9

accuracy		0.6156	398	
macro avg	0.5347	0.5537	0.4918	398
weighted avg	0.6572	0.6156	0.6032	398

Balanced Accuracy (Final Soft-Gated, aggregated): 0.553684692107961

===== Accuracies =====

Per-fold balanced accuracy:

Stage 1: [0.717948717948718, 0.9675324675324675, 0.9675324675324675, 0.974025974025974, 1.0]

Stage 2: [0.8110671936758893, 0.8796296296296297, 0.8376572327044025, 0.8223270440251573, 0.8396226415094339]

Stage 3: [0.5431318681318681, 0.5059523809523809, 0.7580882352941176, 0.3678571428571429, 0.6238095238095238]

Mean per-fold balanced accuracy: [np.float64(0.9254079254079253), np.float64(0.8380607483089026), np.float64(0.5597678302090066)]

Aggregated balanced accuracy:

Stage 1: 0.9337219343696028

Stage 2: 0.8378762964836832

Stage 3: 0.571354264807487

Final Soft-Gated: 0.553684692107961

===== CV RUN 5/5 =====

Starting 5-split CV (3 train / 1 val / 1 test) on 398 samples...

===== Fold 1 =====

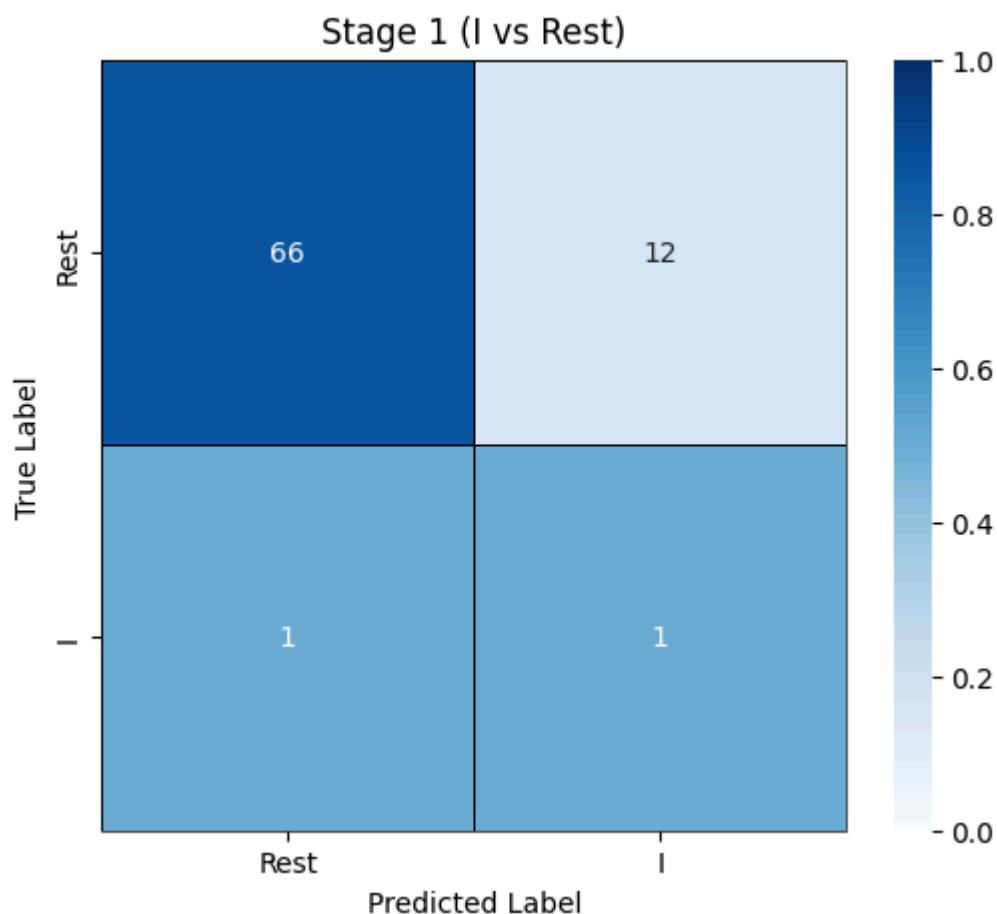
Train: 238, Val: 80, Test: 80

These features will be dropped:

['C11', 'C101', 'C64', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']

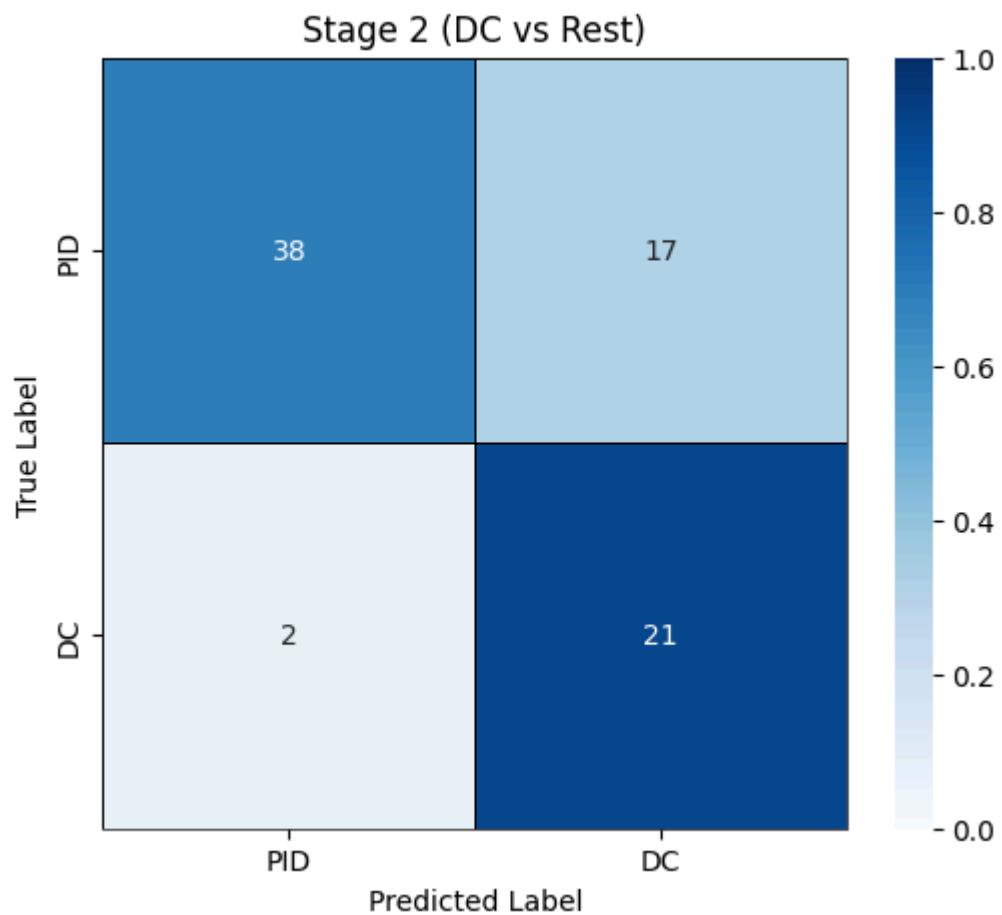
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 0.9935 at threshold=0.648



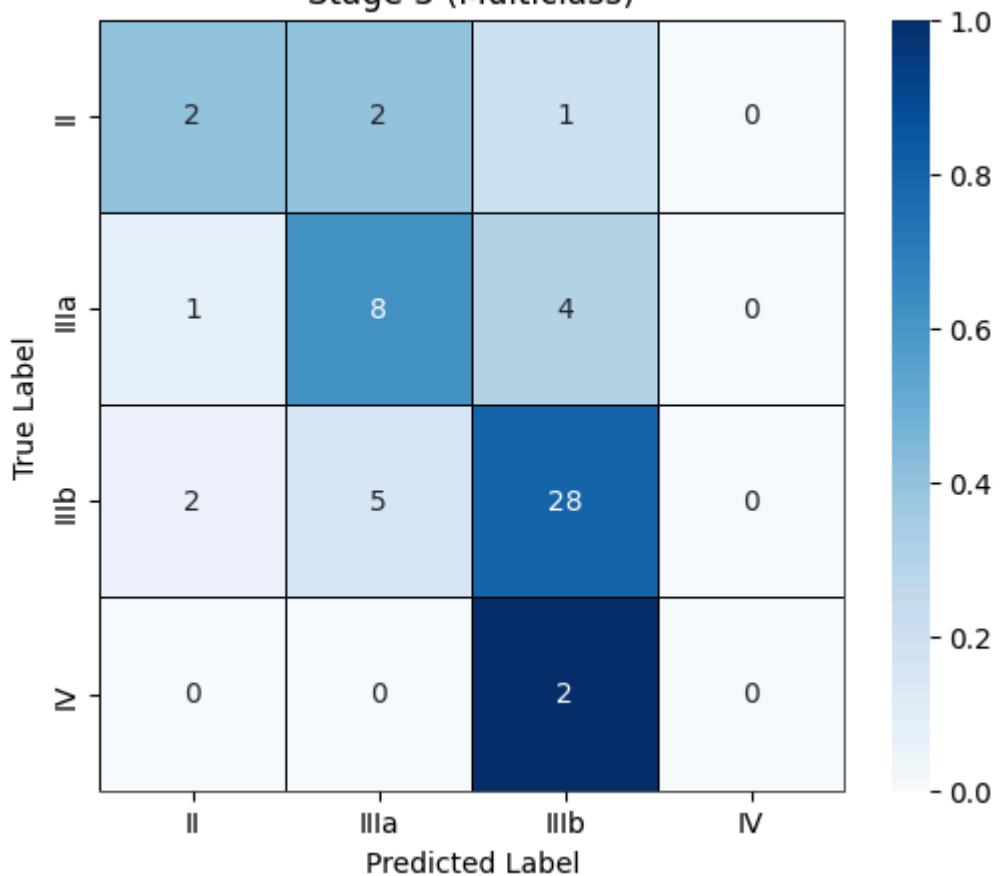
-- Stage 2 (DC vs Rest) --

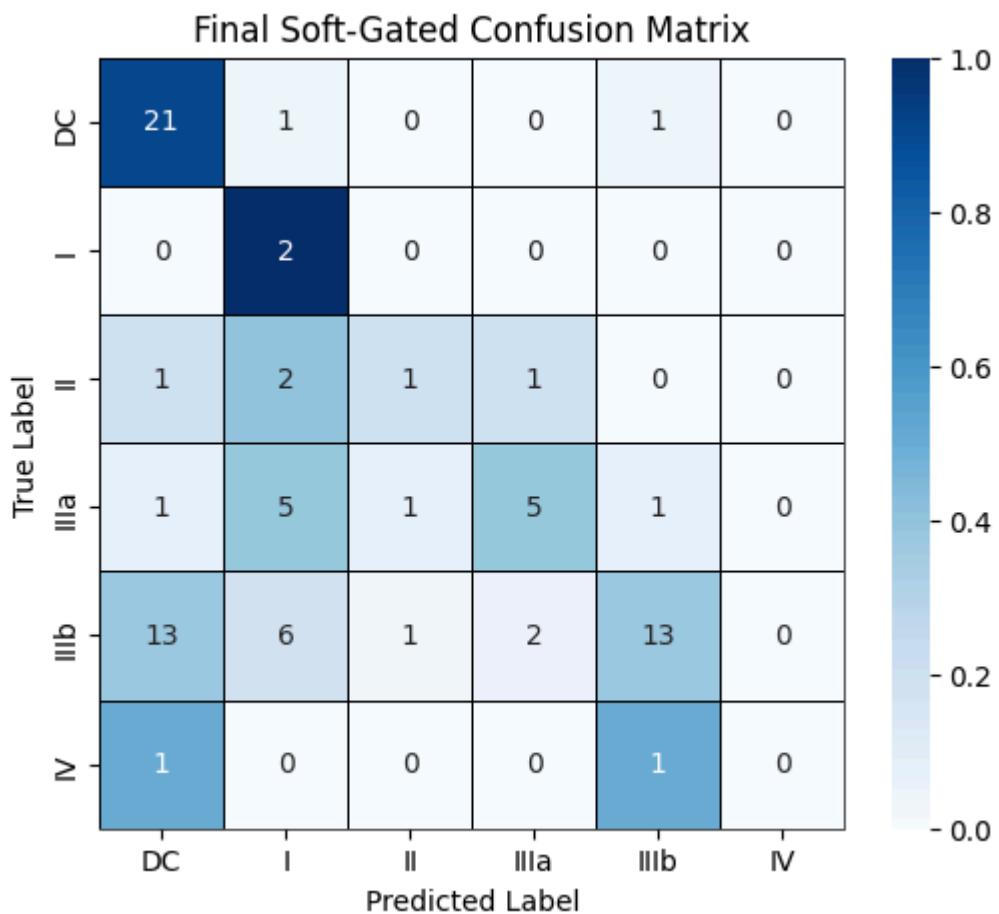
[Threshold Optimization] Best balanced_accuracy: 0.8575 at threshold=0.271



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 2 =====

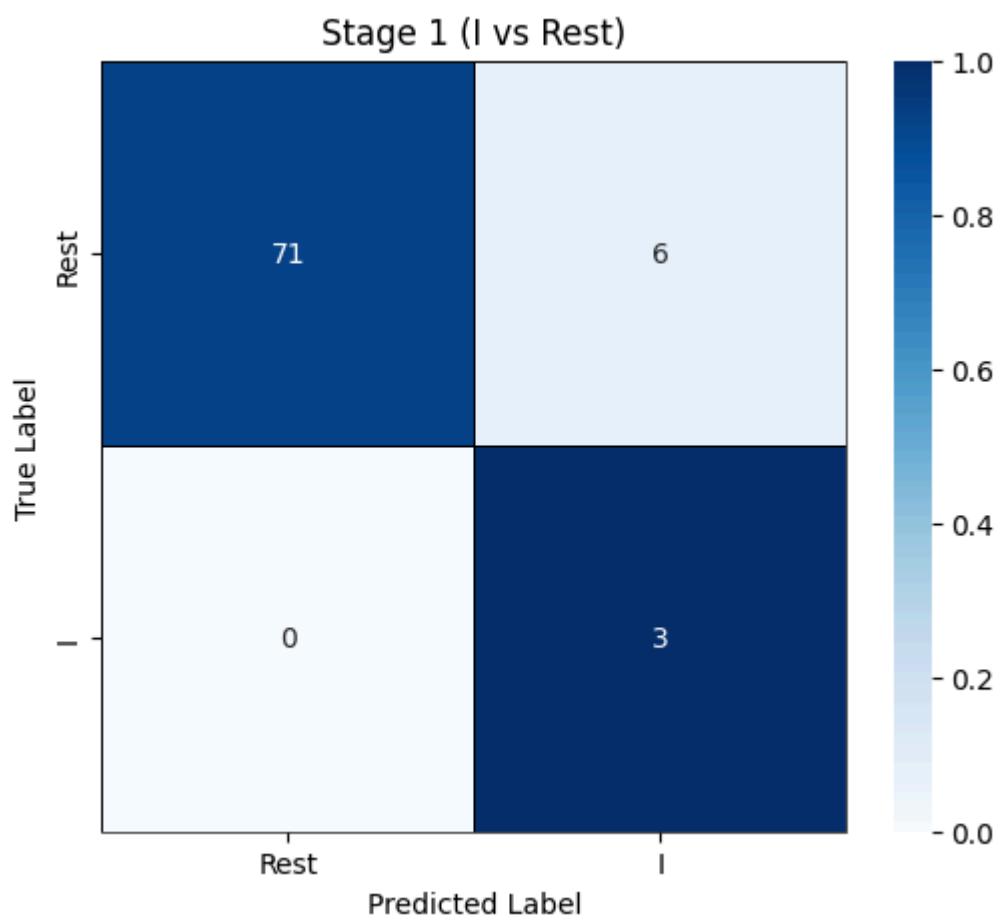
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

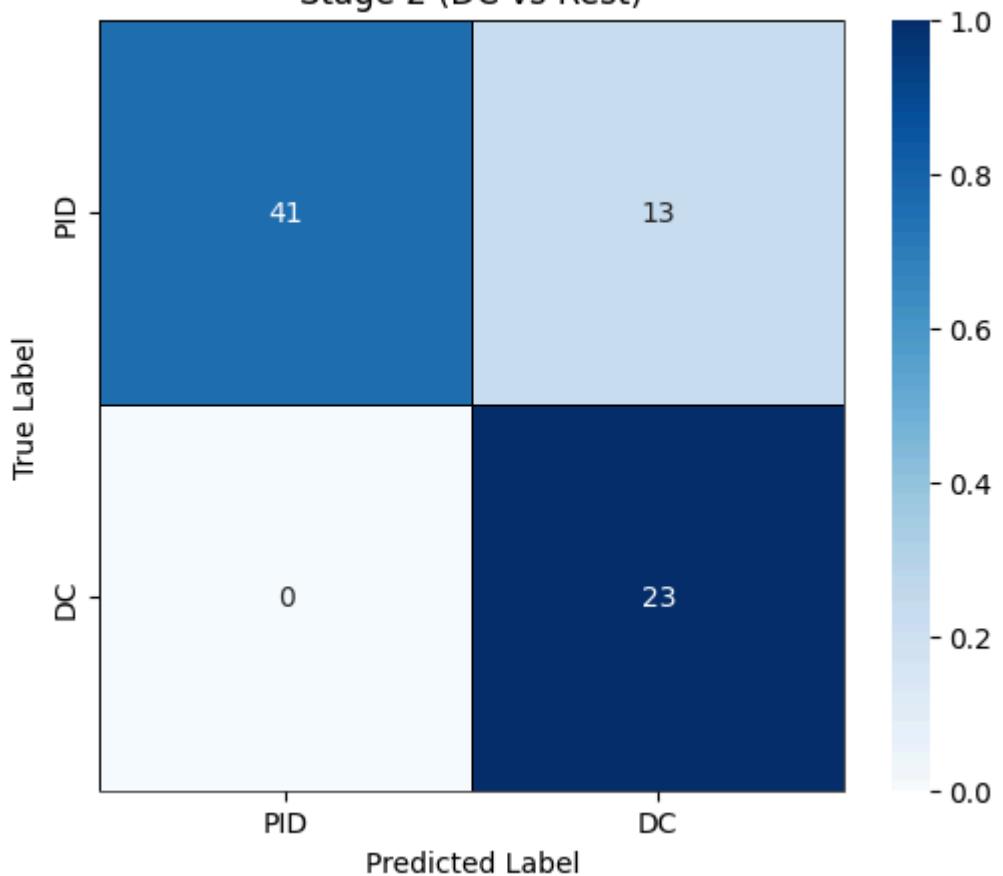
[Threshold Optimization] Best balanced_accuracy: 0.9936 at threshold=0.588



-- Stage 2 (DC vs Rest) --

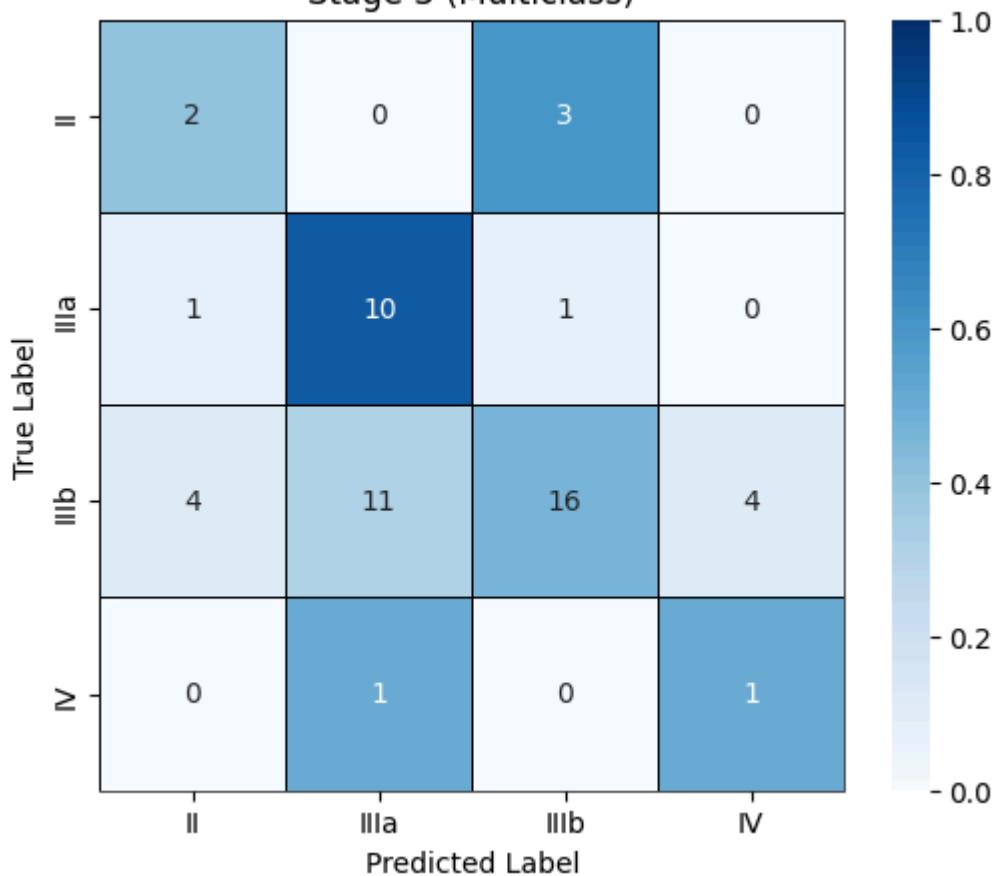
[Threshold Optimization] Best balanced_accuracy: 0.8565 at threshold=0.201

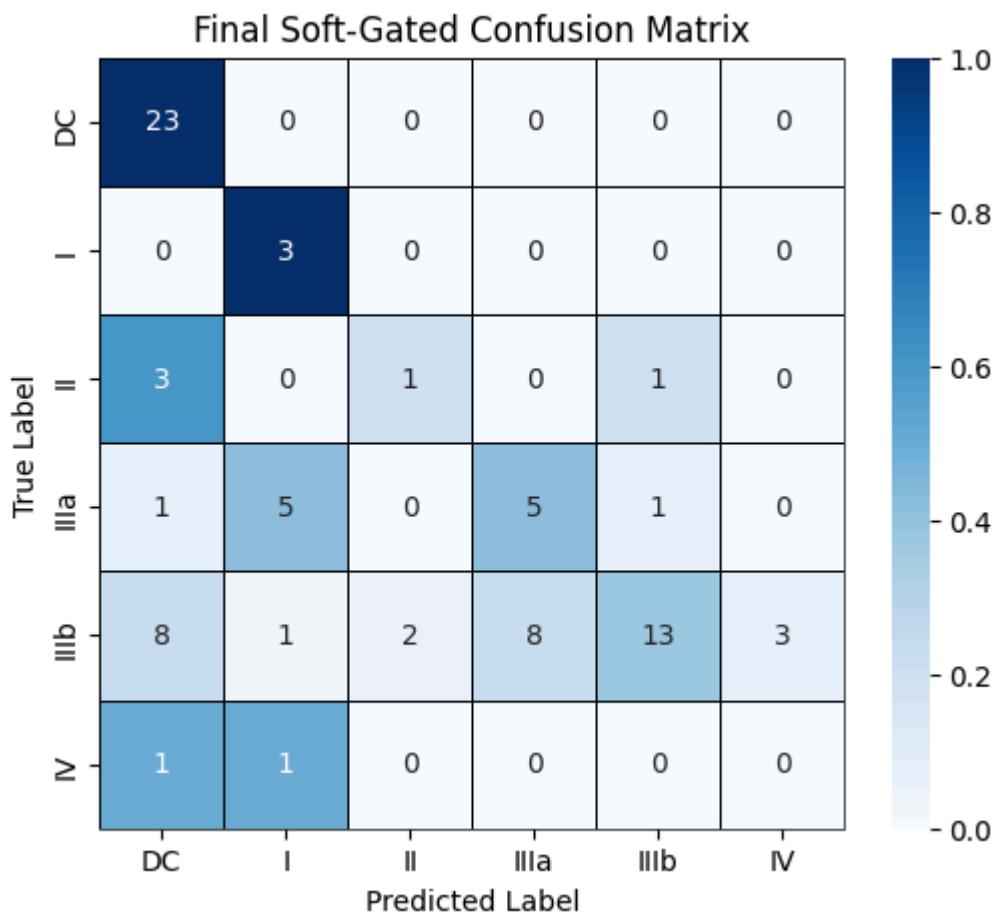
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 3 =====

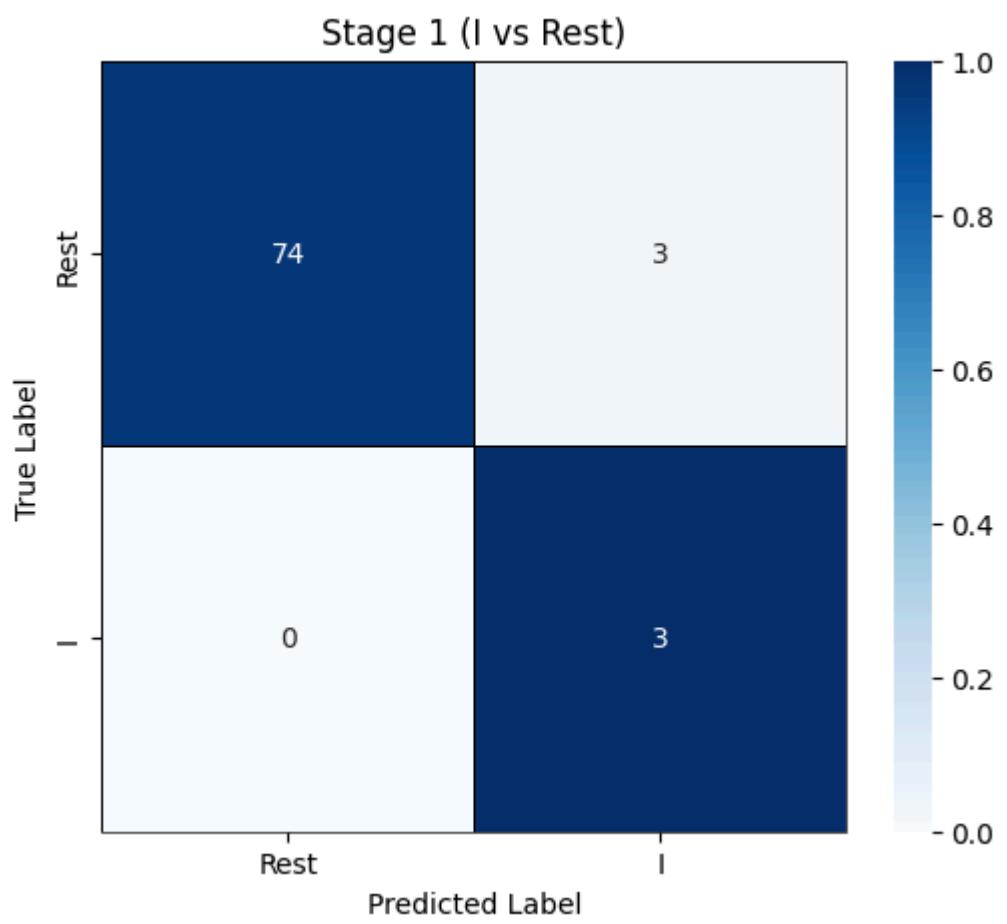
Train: 238, Val: 80, Test: 80

These features will be dropped:

`['C11', 'C101', 'MC6. DP T cells', 'C64', 'C100', 'C126', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

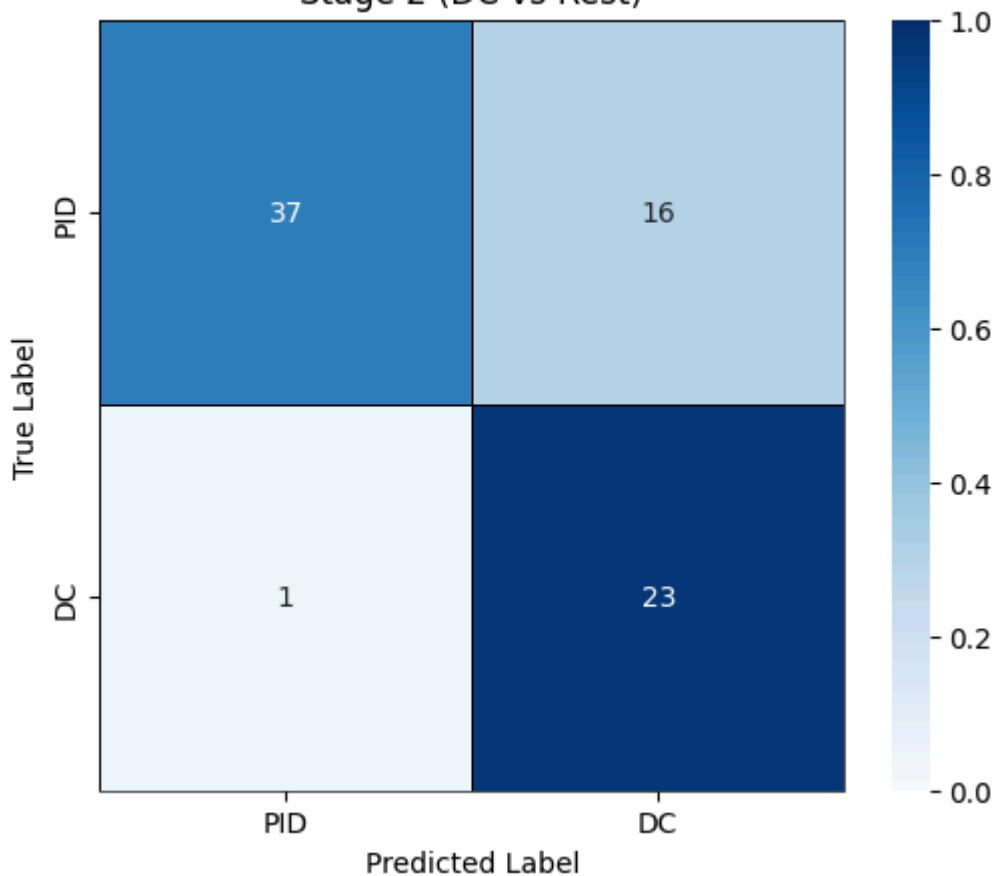
[Threshold Optimization] Best balanced_accuracy: 0.9808 at threshold=0.447



-- Stage 2 (DC vs Rest) --

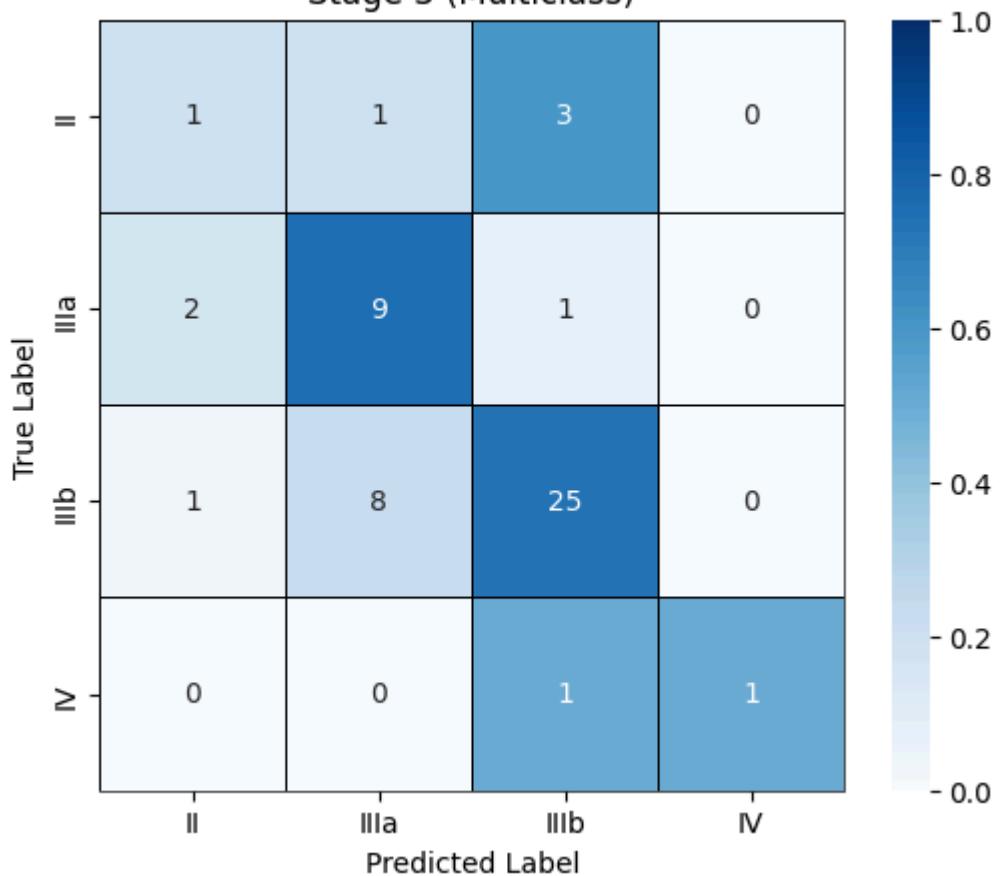
[Threshold Optimization] Best balanced_accuracy: 0.8055 at threshold=0.065

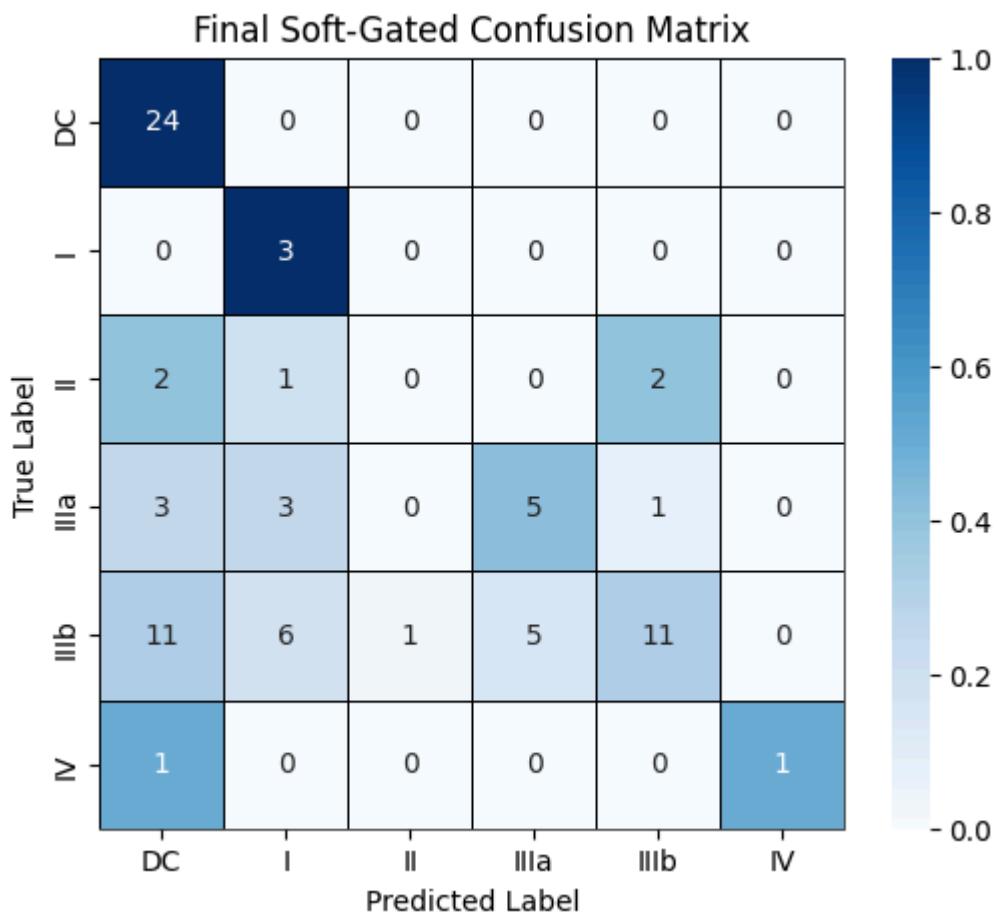
Stage 2 (DC vs Rest)



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 4 =====

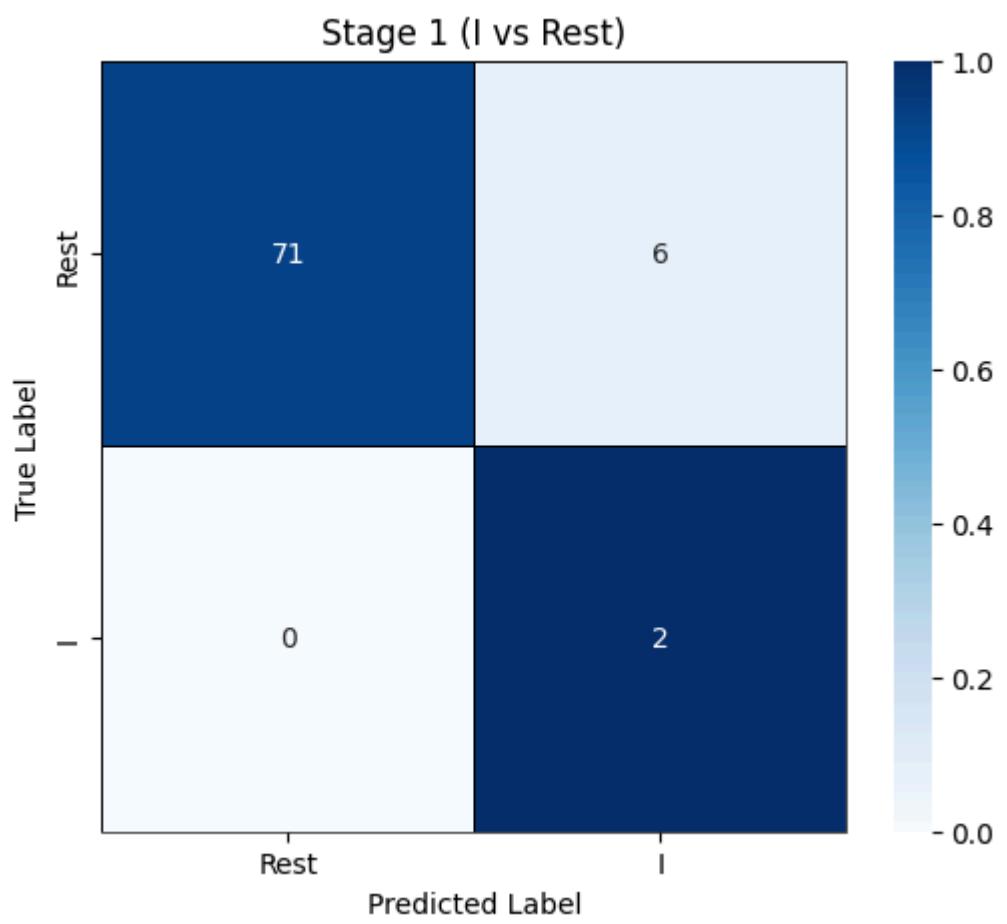
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'C96', 'MC19. TD CD4+ T cells', 'C127', 'C51', 'C135', 'C66']`

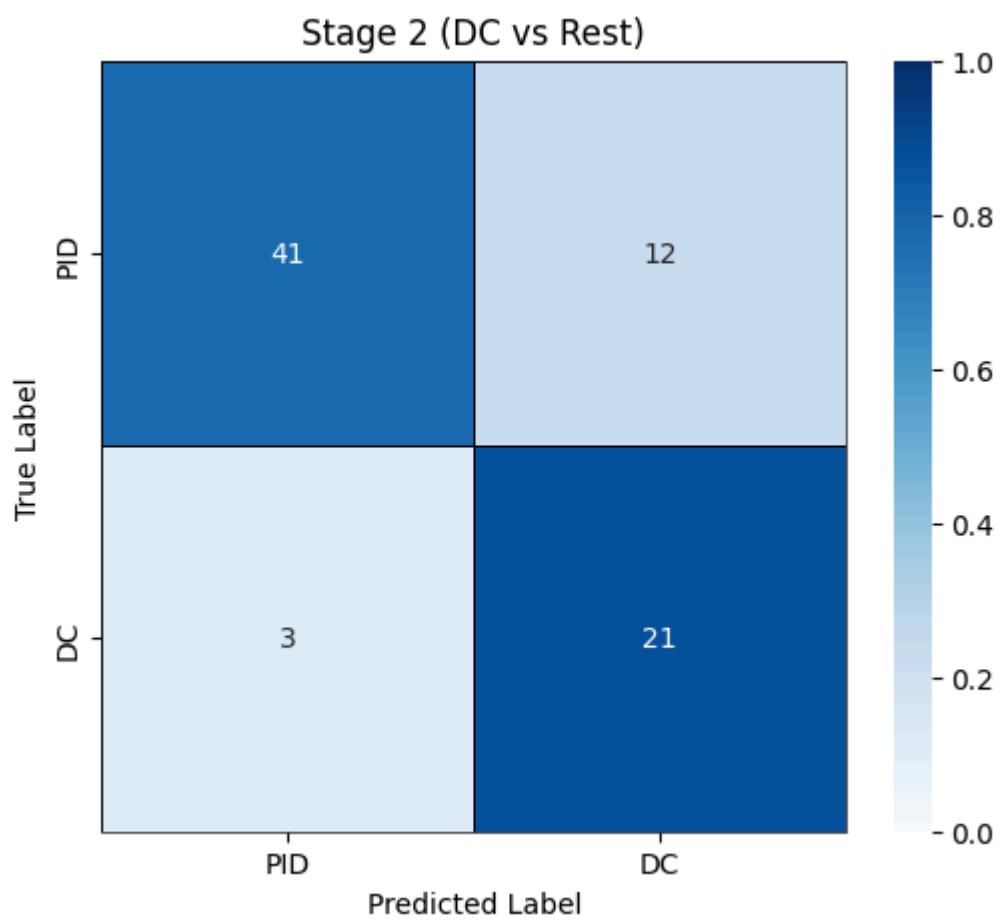
-- Stage 1 (I vs Rest) --

[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.548



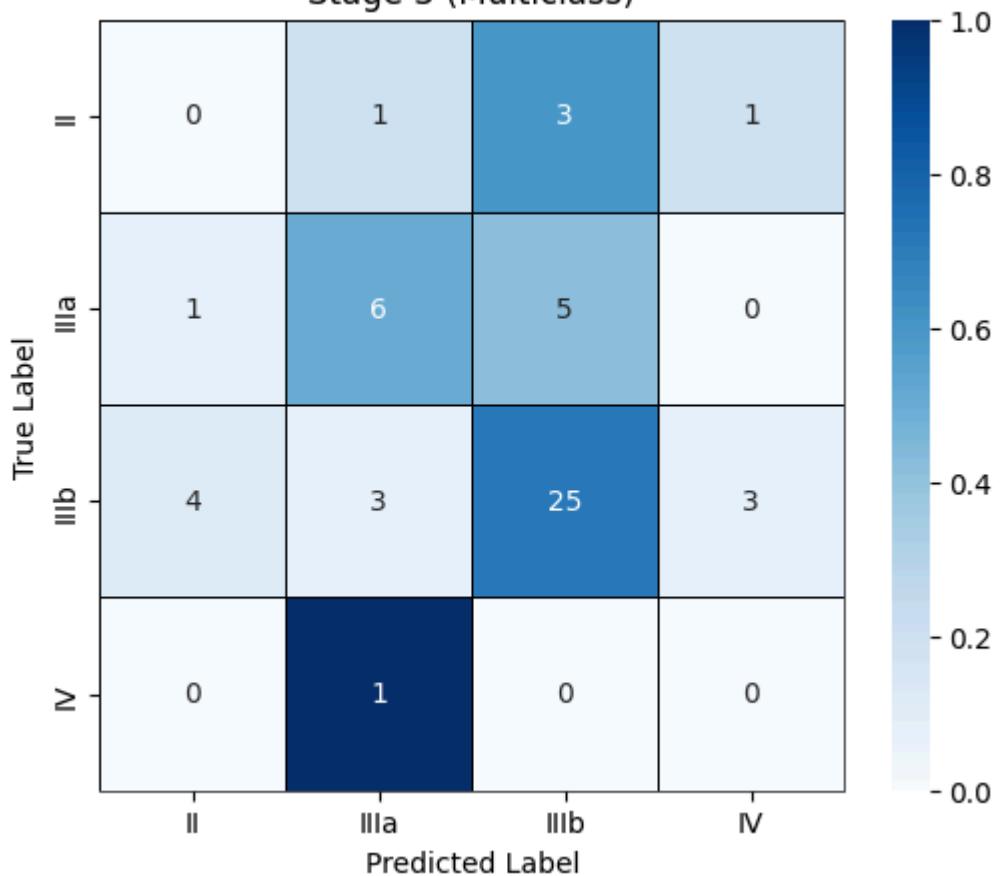
-- Stage 2 (DC vs Rest) --

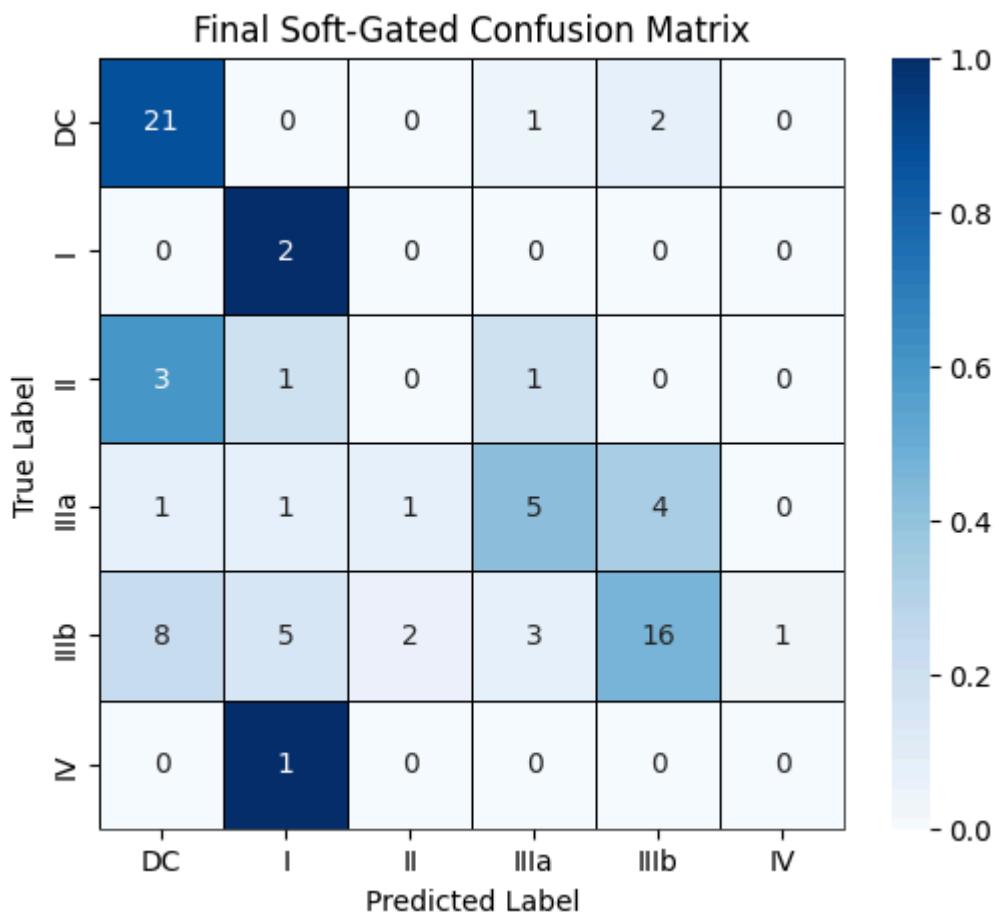
[Threshold Optimization] Best balanced_accuracy: 0.8474 at threshold=0.226



-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)





===== Fold 5 =====

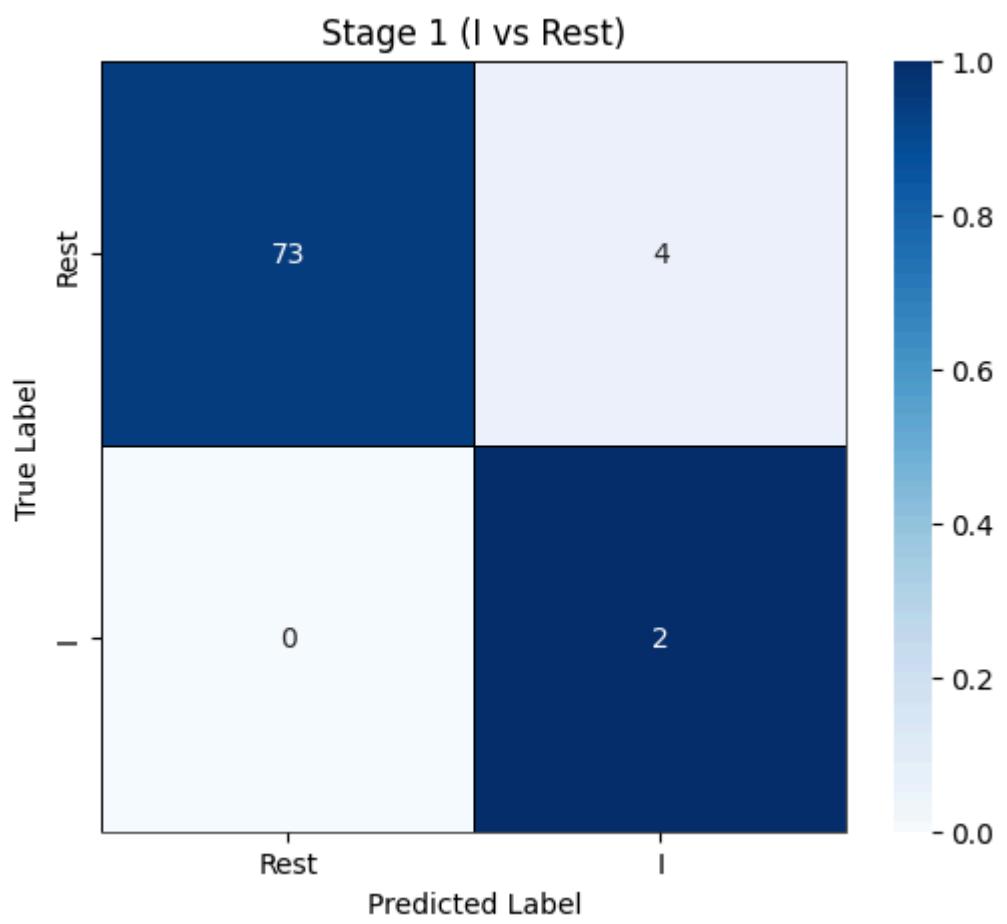
Train: 239, Val: 80, Test: 79

These features will be dropped:

`['C11', 'C101', 'C64', 'C104', 'C100', 'C126', 'MC30. DN T cells', 'C82', 'C139', 'C85', 'C113', 'C87', 'C40', 'C86', 'C92', 'MC28. PreGC B cells', 'C96', 'C127', 'C51', 'C135', 'C66']`

-- Stage 1 (I vs Rest) --

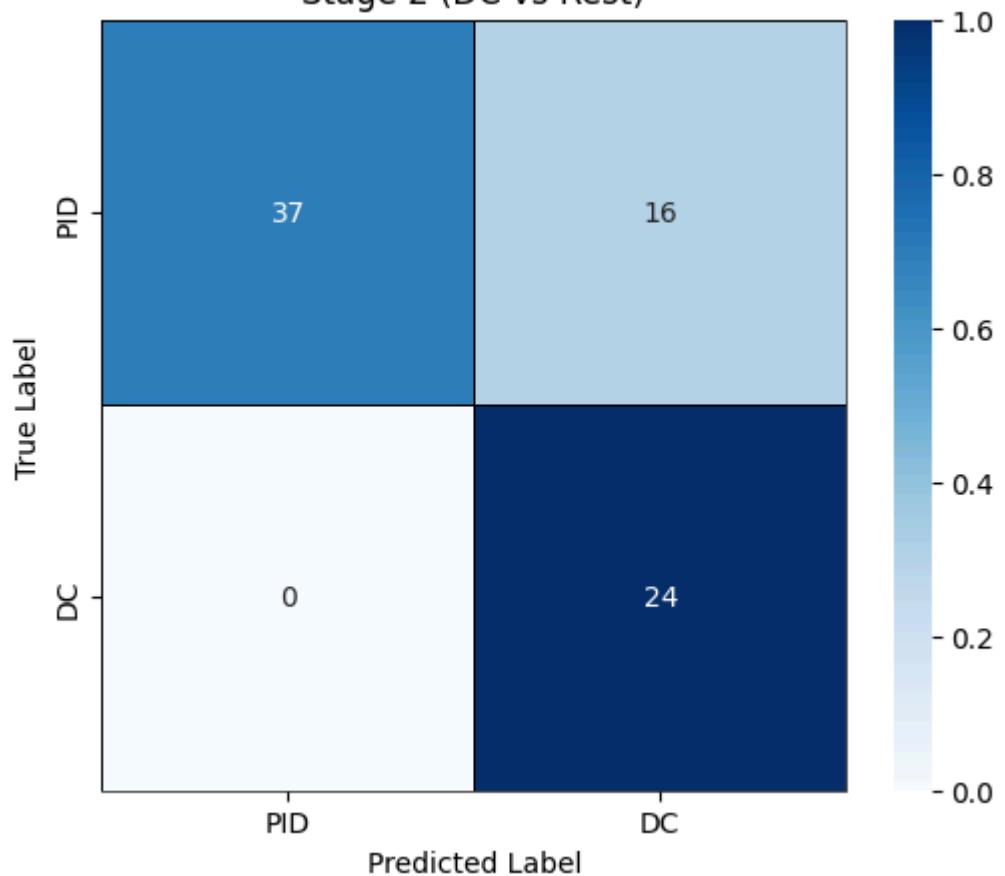
[Threshold Optimization] Best balanced_accuracy: 1.0000 at threshold=0.523



-- Stage 2 (DC vs Rest) --

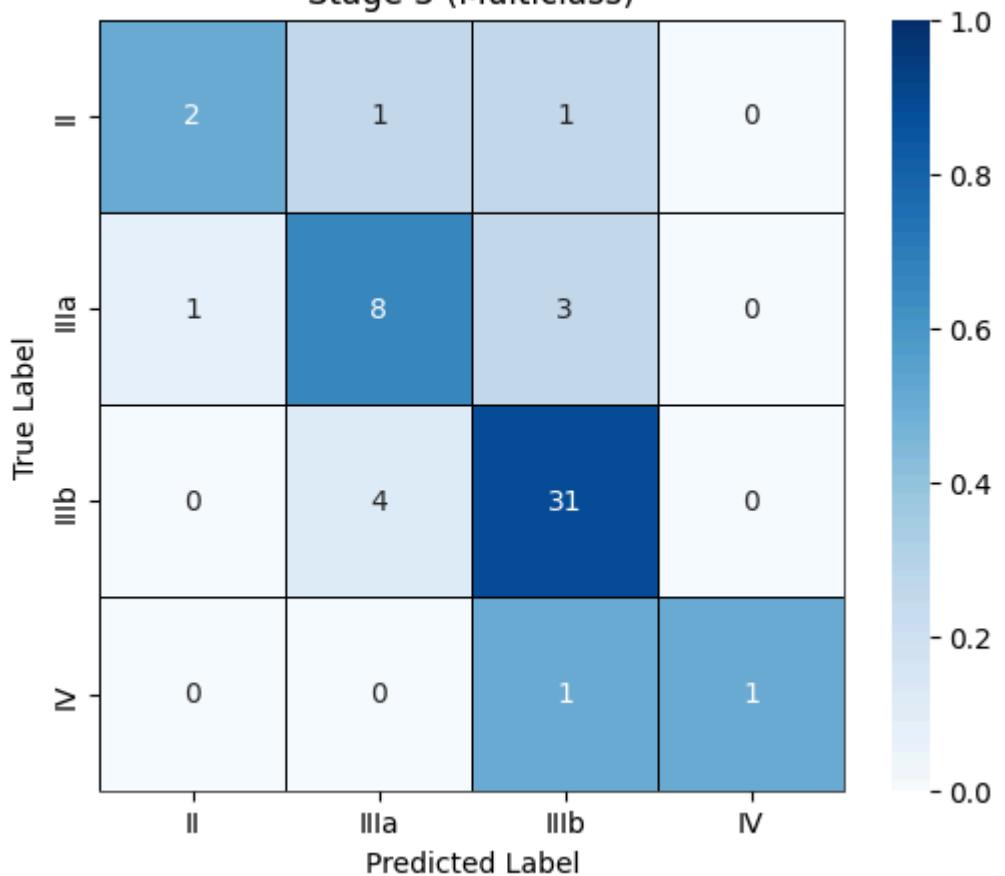
[Threshold Optimization] Best balanced_accuracy: 0.8747 at threshold=0.191

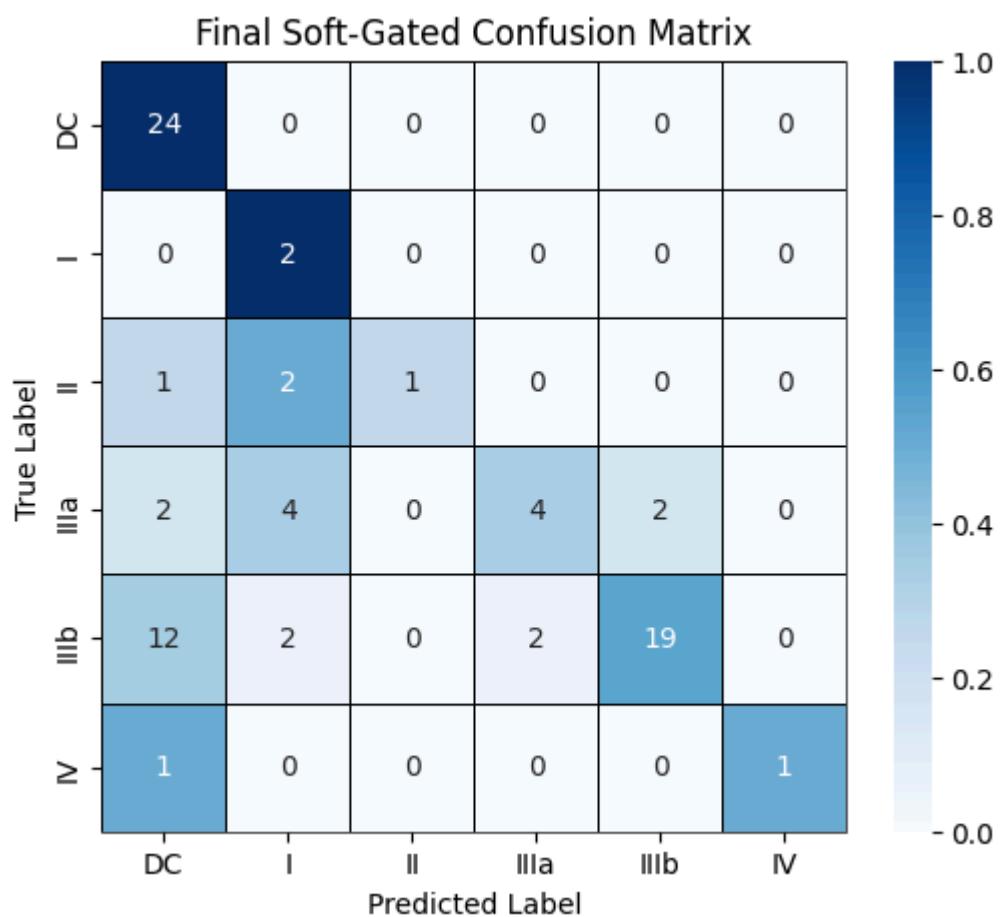
Stage 2 (DC vs Rest)



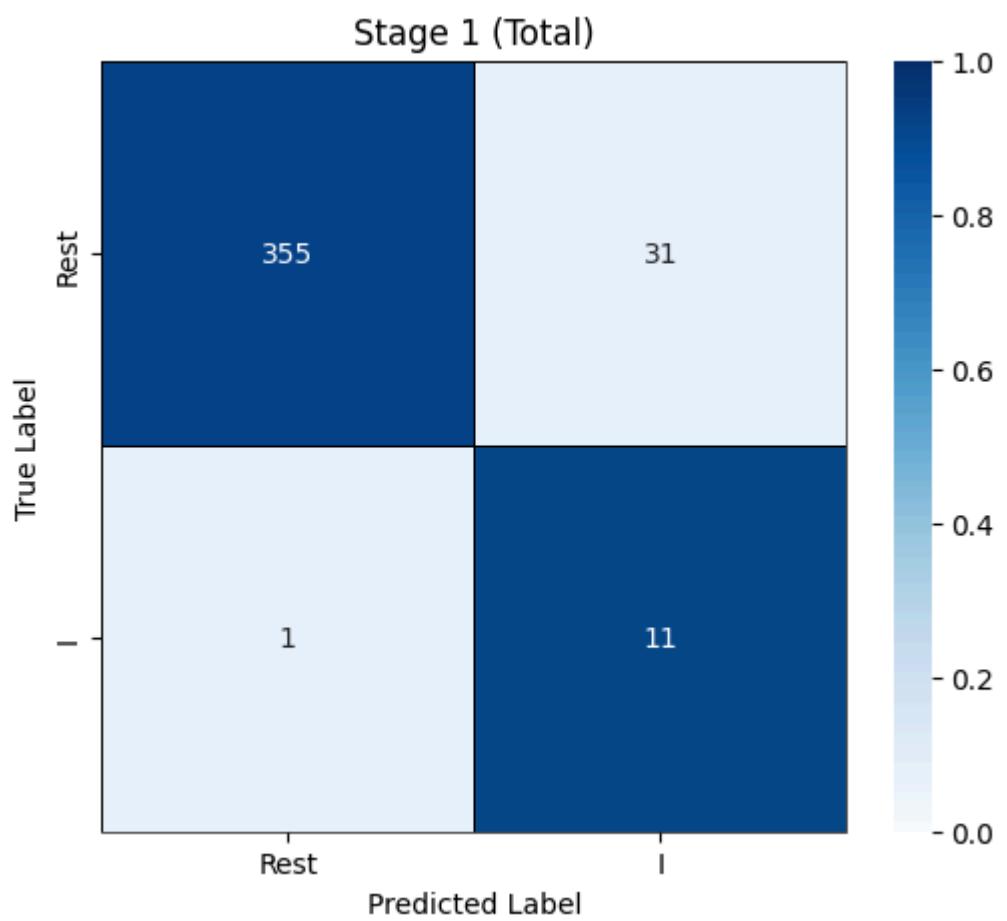
-- Stage 3 (Multiclass) --

Stage 3 (Multiclass)

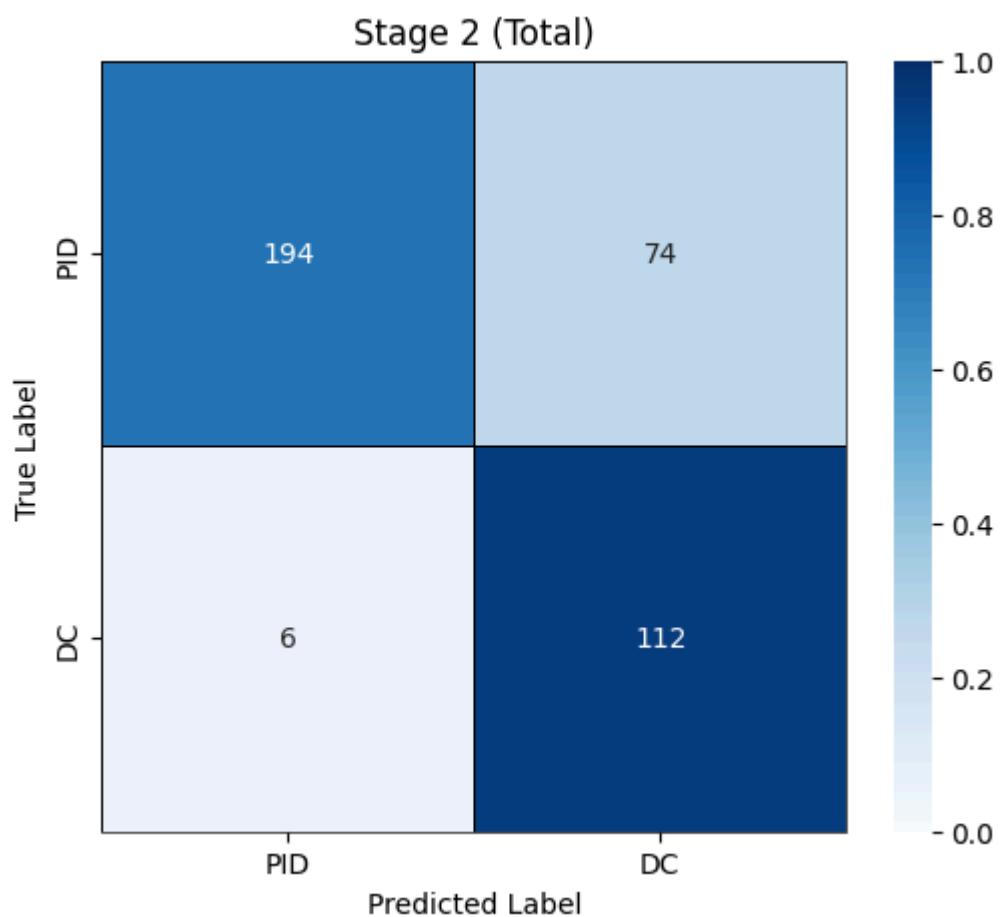




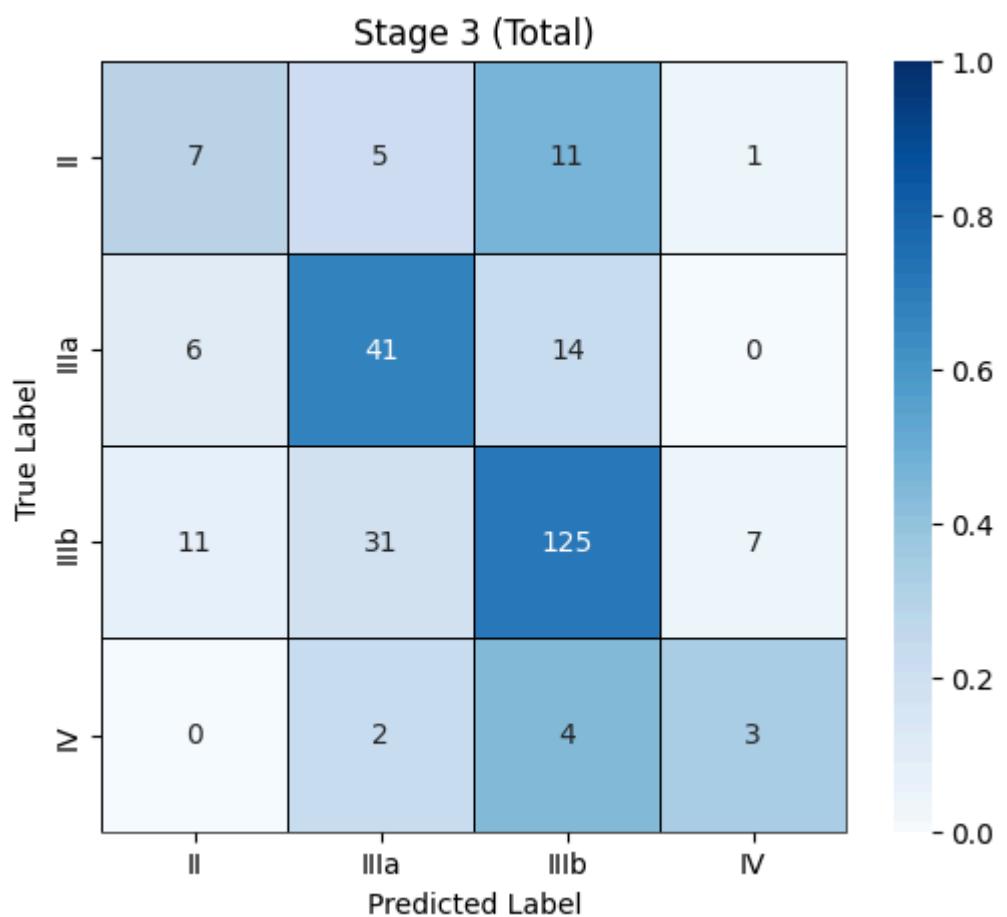
===== AGGREGATED CONFUSION MATRICES =====



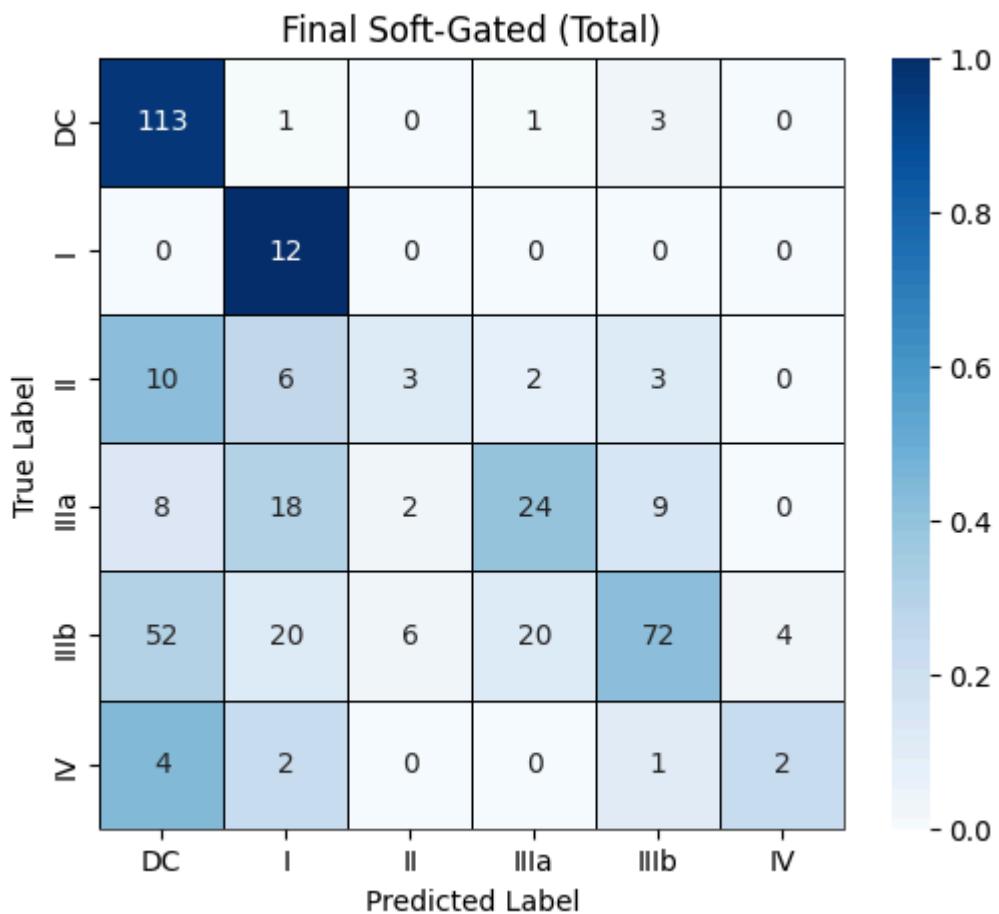
	precision	recall	f1-score	support
0	0.997	0.920	0.957	386
1	0.262	0.917	0.407	12
accuracy		0.920	0.920	398
macro avg	0.630	0.918	0.682	398
weighted avg	0.975	0.920	0.940	398



	precision	recall	f1-score	support
0	0.970	0.724	0.829	268
1	0.602	0.949	0.737	118
accuracy		0.793	0.793	386
macro avg	0.786	0.837	0.783	386
weighted avg	0.858	0.793	0.801	386



	precision	recall	f1-score	support
0	0.292	0.292	0.292	24
1	0.519	0.672	0.586	61
2	0.812	0.718	0.762	174
3	0.273	0.333	0.300	9
accuracy		0.657	0.657	268
macro avg	0.474	0.504	0.485	268
weighted avg	0.680	0.657	0.664	268



===== CV SUMMARY =====

precision recall f1-score support

DC	0.6043	0.9576	0.7410	118
I	0.2034	1.0000	0.3380	12
II	0.2727	0.1250	0.1714	24
IIIa	0.5106	0.3934	0.4444	61
IIIb	0.8182	0.4138	0.5496	174
IV	0.3333	0.2222	0.2667	9

accuracy		0.5678	398	
macro avg	0.4571	0.5187	0.4185	398
weighted avg	0.6452	0.5678	0.5547	398

Balanced Accuracy (Final Soft-Gated, aggregated): 0.5186808445442309

===== Accuracies =====

Per-fold balanced accuracy:

Stage 1: [0.6730769230769231, 0.961038961038961, 0.9805194805194806, 0.961038961038961, 0.974025974025974]

Stage 2: [0.8019762845849803, 0.8796296296296297, 0.8282232704402516, 0.8242924528301887, 0.8490566037735849]

Stage 3: [0.45384615384615384, 0.5476190476190477, 0.5463235294117648, 0.3035714285714286, 0.638095238095238]
Mean per-fold balanced accuracy: [np.float64(0.90994005994006), np.float64(0.836635648251727), np.float64(0.49789107950872663)]
Aggregated balanced accuracy:
Stage 1: 0.9181778929188256
Stage 2: 0.8365165696939034
Stage 3: 0.5038804880346712
Final Soft-Gated: 0.5186808445442309

===== MULTI-RUN SUMMARY =====

Mean per-fold balanced accuracy (across runs):
Stage 1: mean = 0.890 , se = 0.032
Stage 2: mean = 0.838 , se = 0.002
Stage 3: mean = 0.523 , se = 0.013

Aggregated balanced accuracy (across runs):
Stage 1: mean = 0.899 , se = 0.013
Stage 2: mean = 0.838 , se = 0.002
Stage 3: mean = 0.530 , se = 0.012
Final Soft-Gated: mean = 0.517 , se = 0.012

```
[{'mean_per_fold': [0.8956376956376957, 0.8460725176291213, 0.5421536306830423], 'aggregated': [0.9039291882556131, 0.8453389830508475, 0.5505111173921237, 0.5364127061689125]}, {'mean_per_fold': [0.888128538128538, 0.8304388774261622, 0.502059362206421], 'aggregated': [0.8881692573402418, 0.829907665064508, 0.5061652063312606, 0.48538863942268623]}, {'mean_per_fold': [0.8325840825840825, 0.8368168076720168, 0.5128512174100409], 'aggregated': [0.8493091537132988, 0.8365165696939034, 0.5165171000565292, 0.4910313706411282]}, {'mean_per_fold': [0.9254079254079253, 0.8380607483089026, 0.5597678302090066], 'aggregated': [0.9337219343696028,
```

```
0.8378762964836832,  
0.571354264807487,  
0.553684692107961]},  
{'mean_per_fold': [0.90994005994006,  
0.836635648251727,  
0.49789107950872663],  
'aggregated': [0.9181778929188256,  
0.8365165696939034,  
0.5038804880346712,  
0.5186808445442309]}]
```

```
params1 = {'units': 239, 'learning_rate': 0.0016, 'dropout_rate': 0.364, 'num_hidden_layers': 1,  
'batch_size': 110, 'epochs': 46, 'l2_reg': 0.0039, 'activation': "selu", 'optimizer': "sgd"}  
params2 = {'units': 167, 'learning_rate': 0.0054, 'dropout_rate': 0.3488, 'num_hidden_layers': 1,  
'batch_size': 21, 'epochs': 50, 'l2_reg': 0.00612, 'activation': 'selu', 'optimizer': 'rmsprop'}  
params3 = {'units': 86, 'learning_rate': 0.01, 'dropout_rate': 0, 'num_hidden_layers': 3,  
'batch_size': 71, 'epochs': 86, 'l2_reg': 0.000001, 'activation': "selu", 'optimizer': "rmsprop"}
```