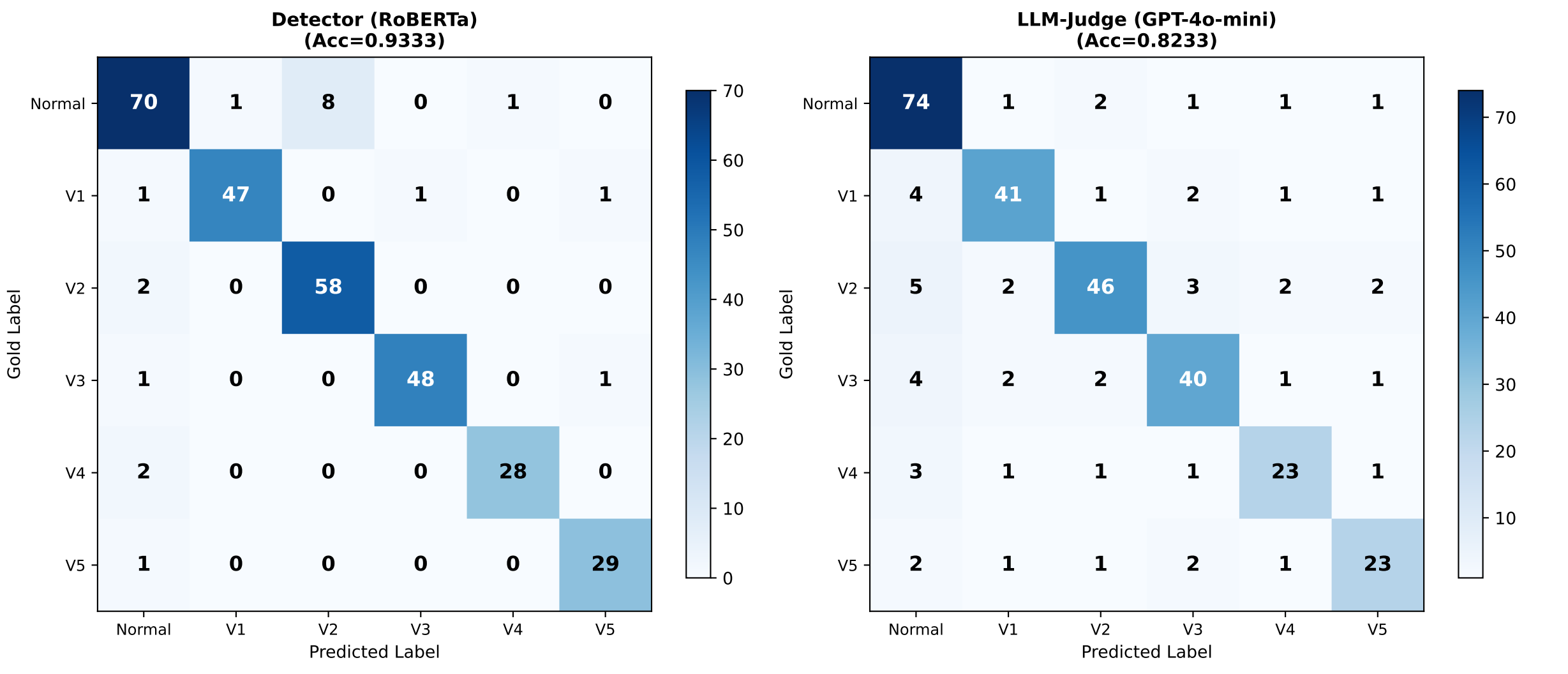
Detector (RoBERTa)
(Acc=0.9333)
Normal
V1
V2
V3
V4
V5
Gold Label
Predicted Label
70 1 8 0 1 0
1 47 0 1 0 1
2 0 58 0 0 0
1 0 0 48 0 1
2 0 0 0 28 0
1 0 0 0 0 29
70
60
50
40
30
20
10
0
LLM-Judge (GPT-4o-mini)
(Acc=0.8233)
74 1 2 1 1 1
4 41 1 2 1 1
5 2 46 3 2 2
4 2 2 40 1 1
3 1 1 1 23 1
2 1 1 2 1 23