Results

May 9, 2024

1 Tables of Friedman, Bonferroni-Dunn, Holm, Hochberg and Hommel Tests

Table 1: Average Rankings of the algorithms

Algorithm	Ranking
CRSPT	2.38
—DeepMARL-MR Prioritized (S)-model-5	3.59
—DeepMARL-MR (S)-model-4	4.21
SPT	4.81
—DeepMARL-AS (8 rules) (GDT)-model-0	5.24
—DeepMARL-AS (Beefy)-model-7	6.49
—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	6.73
LWRK	7.17
—DeepMARL-AS (8 rules) Prioritized (S)-model-3	8.11
—DeepMARL-AS (8 rules) (S)-model-1	8.58
MS	10.09
WINQ	10.62

Friedman statistic considering reduction performance (distributed according to chi-square with 11 degrees of freedom: 1671.4728205128083.
P-value computed by Friedman Test: 0.0.

Iman and Davenport statistic considering reduction performance (distributed according to F-distribution with 11 and 3289 degrees of freedom: 306.88488324198727.
P-value computed by Iman and Daveport Test: 2.220446049250313E-16.

Table 2: Holm / Hochberg Table for $\alpha = 0.05$

i	algorithm	$z = (R_0 - R_i)/SE$	p	Holm/Hochberg/Hommel
11	WINQ	27.97856551371287	2.962603460225532E-172	0.004545454545454546
10	MS	26.189567799764497	3.493880774457326E-151	0.005
9	—DeepMARL-AS (8 rules) (S)-model-1	21.07167560543091	1.447187429074700 4E-98	0.005555555555555556
8	—DeepMARL-AS (8 rules) Prioritized (S)-model-3	19.452519446604125	2.7745761202980874E-84	0.00625
7	LWRK	16.259498210316366	1.9129822757584764E-59	0.007142857142857143 5
6	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	14.77621529558693	2.085609981452213E-49	0.008333333333333333
5	—DeepMARL-AS (Beefy)-model-7	13.949653060661385	3.1613947775814485E-44	0.01
4	—DeepMARL-AS (8 rules) (GDT)-model-0	9.714936952960603	2.6041013258025245E-22	0.0125
3	SPT	8.24297680857262	1.6797814098869758E-16	0.016666666666666666
2	—DeepMARL-MR (S)-model-4	6.204878147112357	5.473929939087479E-10	0.025
1	—DeepMARL-MR Prioritized (S)-model-5	4.098842863603422	4.1522080403999844E-5	0.05

Bonferroni-Dunn's procedure rejects those hypotheses that have a p-value $\leq 0.004545454545454546$.
Hochberg's procedure rejects those hypotheses that have a p-value ≤ 0.05.
Hommel's procedure rejects all hypotheses.

Table 3: Holm / Hochberg Table for $\alpha = 0.10$

i	algorithm	$z = (R_0 - R_i)/SE$	p	Holm/Hochberg/Hommel
11	WINQ	27.97856551371287	2.962603460225532E-172	0.009090909090909092
10	MS	26.189567799764497	3.493880774457326E-151	0.01
9	—DeepMARL-AS (8 rules) (S)-model-1	21.07167560543091	1.4471874290747004E-98	0.011111111111111112
8	—DeepMARL-AS (8 rules) Prioritized (S)-model-3	19.452519446604125	2.7745761202980874E-84	0.0125
7	LWRK	16.259498210316366	1.9129822757584764E-59	0.014285714285714287
6	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	14.77621529558693	2.085609981452213E-49	0.016666666666666666
5	—DeepMARL-AS (Beefy)-model-7	13.949653060661385	3.1613947775814485E-44	0.02
4	—DeepMARL-AS (8 rules) (GDT)-model-0	9.714936952960603	2.6041013258025245E-22	0.025
3	SPT	8.24297680857262	1.6797814098869758E-16	0.03333333333333333
2	—DeepMARL-MR (S)-model-4	6.20487814712357	5.473929939087479E-10	0.05
1	—DeepMARL-MR Prioritized (S)-model-5	4.098842863603422	4.1522080403999844E-5	0.1

Bonferroni-Dunn's procedure rejects those hypotheses that have a p-value $\leq 0.009090909090909092$.
Hochberg's procedure rejects those hypotheses that have a p-value ≤ 0.1.
Hommel's procedure rejects all hypotheses.

Table 4: Adjusted p-values

i	algorithm	unadjusted p	p_{Bonf}	p_{Holm}	p_{Hoch}	p_{Homm}
1	WINQ	2.962603460225532E-172	3.258863806248085E-171	3.258863806248085E-171	3.258863806248085E-171	3.258863806248085E-171
2	MS	3.493880774457326E-151	3.8432688519030586E-150	3.4938807744573264E-150	3.4938807744573264E-150	3.4938807744573264E-150
3	—DeepMARL-AS (8 rules) (S)-model-1	1.4471874290747004E-98	1.5919061719821706E-97	1.3024686861672305E-97	1.3024686861672305E-97	1.3024686861672305E-97
4	—DeepMARL-AS (8 rules) Prioritized (S)-model-3	2.7745761202980874E-84	3.0520337323278962E-83	2.21966089623847E-83	2.21966089623847E-83	2.21966089623847E-83
5	LWRK	1.9129822757584764E-59	2.1042805033343242E-58	1.3390875930309335E-58	1.3390875930309335E-58	1.3390875930309335E-58
6	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	2.085609981452213E-49	2.2941709795974345E-48	1.2513659888713278E-48	1.2513659888713278E-48	1.2513659888713278E-48
7	—DeepMARL-AS (Beefy)-model-7	3.1613947775814485E-44	3.4775342553395933E-43	1.5806973887907242E-43	1.5806973887907242E-43	1.5806973887907242E-43
8	—DeepMARL-AS (8 rules) (GDT)-model-0	2.6041013258025245E-22	2.864511458382777E-21	1.0416405303210098E-21	1.0416405303210098E-21	1.0416405303210098E-21
9	SPT	1.6797814098869758E-16	1.8477595508756734E-15	5.039344229660927E-16	5.039344229660927E-16	5.039344229660927E-16
10	—DeepMARL-MR (S)-model-4	5.473929939087479E-10	6.021322932996226 5E-9	1.0947859878174958E-9	1.0947859878174958E-9	1.0947859878174958E-9
11	—DeepMARL-MR Prioritized (S)-model-5	4.1522080403999844E-5	4.567428444399834E-4	4.1522080403999844E-5	4.1522080403999844E-5	4.1522080403999844E-5

Nemenyi's procedure rejects those hypotheses that have a p-value $\leq 7.575757575757576E-4$.
Holm's procedure rejects those hypotheses that have a p-value ≤ 0.00625.
Shaffer's procedure rejects those hypotheses that have a p-value $\leq 7.575757575757576E-4$.
Nemenyi's procedure rejects those hypotheses that have a p-value $\leq 0.0015151515151515152$.
Holm's procedure rejects those hypotheses that have a p-value ≤ 0.0125.
Shaffer's procedure rejects those hypotheses that have a p-value $\leq 0.0015151515151515152$.

Table 5: Holm / Shaffer Table for $\alpha = 0.05$

i	algorithms	$z = (R_0 - R_i)/SE$	p	Holm	Shaffer
66	CRSPT vs. WINQ	27.97856551371287	2.962603460225532E-172	7.575757575757576E-4	7.575757575757576E-4
65	CRSPT vs. MS	26.189567799764497	3.493880774457326E-151	7.692307692307692E-4	9.090909090909091E-4
64	—DeepMARL-MR Prioritized (S)-model-5 vs. WINQ	23.87972265010945	4.9757581244479696E-126	7.8125E-4	9.090909090909091E-4
63	—DeepMARL-MR Prioritized (S)-model-5 vs. MS	22.090724936161074	3.8811750949767075E-108	7.936507936507937E-4	9.090909090909091E-4
62	—DeepMARL-MR (S)-model-4 vs. WINQ	21.773687366600512	4.121067052782224E-105	8.064516129032258E-4	9.090909090909091E-4
61	CRSPT vs. —DeepMARL-AS (8 rules) (S)-model-1	21.07167560543091	1.4471874290747004E-98	8.19672131147541E-4	9.090909090909091E-4
60	—DeepMARL-MR (S)-model-4 vs. MS	19.98468965265214	7.48511854269219E-89	8.333333333333334E-4	9.090909090909091E-4
59	SPT vs. WINQ	19.73558870514025	1.066968205209761E-86	8.474576271186442E-4	9.090909090909091E-4
58	CRSPT vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	19.452519446604125	2.7745761202980874E-84	8.620689655172415E-4	9.090909090909091E-4
57	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. WINQ	18.263628560752267	1.6121263764696985E-74	8.771929824561404E-4	9.090909090909091E-4
56	SPT vs. MS	17.946590991191876	5.102802922815227E-72	8.928571428571429E-4	9.090909090909091E-4
55	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) (S)-model-1	16.972832741827492	1.3048306912228623E-64	9.090909090909091E-4	9.090909090909091E-4
54	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. MS	16.474630846803894	5.582829897620006E-61	9.25925925925926E-4	0.0010869565217391304
53	CRSPT vs. LWRK	16.259498210316366	1.9129822757584764E-59	9.433962264150943E-4	0.0010869565217391304
52	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	15.353676583000704	3.3466951445833262E-53	9.615384615384616E-4	0.0010869565217391304
51	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) (S)-model-1	14.866797458318556	5.414200488257255E-50	9.80392156862745E-4	0.0010869565217391304
50	CRSPT vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	14.77621529558693	2.085609981452213E-49	0.001	0.0010869565217391304
49	—DeepMARL-AS (Beefy)-model-7 vs. WINQ	14.028912453051483	1.0373008508511168E-44	0.0010204081632653062	0.0010869565217391304
48	CRSPT vs. —DeepMARL-AS (Beefy)-model-7	13.949653060661385	3.1613947775814485E-44	0.0010416666666666667	0.0010869565217391304
47	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	13.247641299491768	4.6560796370062935E-40	0.0010638297872340426	0.0010869565217391304
46	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. WINQ	13.202350218125941	8.503983897021281E-40	0.0010869565217391304	0.0010869565217391304
45	SPT vs. —DeepMARL-AS (8 rules) (S)-model-1	12.828698796858296	1.1325019874700095E-37	0.0011111111111111111	0.0011111111111111111
44	—DeepMARL-AS (Beefy)-model-7 vs. MS	12.23991473910311	1.9023236122627718E-34	0.0011363636363636365	0.001282051282051282
43	—DeepMARL-MR Prioritized (S)-model-5 vs. LWRK	12.160655346712945	5.035311126415846E-34	0.0011627906976744186	0.001282051282051282
42	LWRK vs. WINQ	11.719067303396503	1.0178644267536058E-31	0.0011904761904761906	0.001282051282051282
41	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. MS	11.413352504177567	3.586350655302524E-30	0.0012195121951219512	0.001282051282051282
40	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) (S)-model-1	11.35673865247031	6.866023872019803E-30	0.00125	0.001282051282051282
39	SPT vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	11.209542638031506	3.6608007587407133E-29	0.001282051282051282	0.001282051282051282
38	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	10.677372431983507	1.2989630871013547E-26	0.0013157894736842105	0.0013513513513513514
37	—DeepMARL-MR (S)-model-4 vs. LWRK	10.054620063204009	8.765890702799013E-24	0.0013513513513513514	0.0013513513513513514
36	LWRK vs. MS	9.930069589448129	3.080404863646444E-23	0.0013888888888888889	0.0013888888888888889
35	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (Beefy)-model-7	9.850810197057964	6.7993636111102757E-23	0.0014285714285714286	0.0014705882352941176
34	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	9.737582493643522	2.0845432002701058E-22	0.0014705882352941176	0.0014705882352941176
33	CRSPT vs. —DeepMARL-AS (8 rules) (GDT)-model-0	9.714936952960603	2.6041013258025245E-22	0.0015151515151515152	0.0016129032258064516
32	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	8.571337148474571	1.0229029222807898E-17	0.0015625	0.0016129032258064516
31	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. WINQ	8.526046067108744	1.51434974485467E-17	0.0016129032258064516	0.0016129032258064516
30	CRSPT vs. SPT	8.24297680857262	1.679781409886975 8E-16	0.0016666666666666668	0.0016666666666666668
29	SPT vs. LWRK	8.016521401743747	1.0878197848639693E-15	0.00172413793103448 3	0.00172413793103448 3
28	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (Beefy)-model-7	7.744774913549029	9.575131268231217E-15	0.0017857142857142859	0.0017857142857142859
27	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) (S)-model-1	7.122022544769528	1.0635465389391118E-12	0.001851851851851852	0.001851851851851852
26	—DeepMARL-AS (8 rules) (S)-model-1 vs. WINQ	6.906889908281955	4.95394079040604E-12	0.0019230769230769232	0.002
25	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. MS	6.737048353160372	1.6163640399222215E-11	0.002	0.002
24	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. LWRK	6.544561257355763	5.967030269282831E-11	0.0020833333333333333	0.0020833333333333333
23	SPT vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	6.53323848701431	6.436253076989309E-11	0.002173913043478261	0.002173913043478261
22	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. —DeepMARL-AS (8 rules) (S)-model-1	6.295460309843985	3.0648897808150574E-10	0.002272727272727273	0.002272727272727273
21	CRSPT vs. —DeepMARL-MR (S)-model-4	6.204878147112357	5.473929939087479E-10	0.00238095238095238 1	0.00238095238095238 1
20	SPT vs. —DeepMARL-AS (Beefy)-model-7	5.7066762520887675	1.1520360652344736E-8	0.0025	0.0025
19	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) (GDT)-model-0	5.616094089357181	1.9532232889899644E-8	0.002631578947368421	0.002631578947368421
18	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	5.502866385942738	3.736656126280505E-8	0.002777777777777778	0.002777777777777778
17	—DeepMARL-AS (8 rules) (S)-model-1 vs. MS	5.1178921943335824	3.0896911055583667E-7	0.0029411764705882353	0.0029411764705882353
16	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	5.061278342626325	4.164547093898287E-7	0.003125	0.003125
15	LWRK vs. —DeepMARL-AS (8 rules) (S)-model-1	4.8121773951145475	1.4929474357434183E-6	0.0033333333333333335	0.0033333333333333335
14	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	4.676304151017196	2.9209129269717458E-6	0.0035714285714285718	0.0035714285714285718
13	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (Beefy)-model-7	4.234716107700783	2.2884062091512816E-5	0.0038461538461538464	0.0038461538461538464
12	—DeepMARL-MR Prioritized (S)-model-5 vs. SPT	4.144133944969196	3.41100338379446E-5	0.004166666666666667	0.004166666666666667
11	CRSPT vs. —DeepMARL-MR Prioritized (S)-model-5	4.098842863603422	4.1522080403999844E-5	0.004545454545454546	0.004545454545454546
10	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) (GDT)-model-0	3.5100588058482463	4.4800760165300123E-4	0.005	0.005
9	LWRK vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	3.193021236287758	0.001407925919310537	0.005555555555555556	0.005555555555555556
8	—DeepMARL-AS (Beefy)-model-7 vs. LWRK	2.30984514965498	0.020896728657797497	0.00625	0.00625
7	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-MR (S)-model-4	2.106035283508935	0.035201287589215	0.007142857142857143 5	0.007142857142857143 5
6	—DeepMARL-MR (S)-model-4 vs. SPT	2.038098661460262	0.04154006700988629	0.008333333333333333	0.008333333333333333
5	MS vs. WINQ	1.7889977139483728	0.07361518372016992	0.01	0.01
4	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. —DeepMARL-AS (8 rules) (S)-model-1	1.6191561588267895	0.10541366817859567	0.0125	0.0125
3	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. LWRK	1.483282914729438	0.13799925973447852	0.016666666666666666	0.016666666666666666
2	SPT vs. —DeepMARL-AS (8 rules) (GDT)-model-0	1.4719601443879842	0.1410316405207138	0.025	0.025
1	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	0.8265622349255423	0.4084852298799626	0.05	0.05

Table 6: Holm / Shaffer Table for $\alpha = 0.10$

i	algorithms	$z = (R_0 - R_i)/SE$	p	Holm	Shaffer
66	CRSPT vs. WINQ	27.97856551371287	2.962603460225532E-172	0.0015151515151515152	0.0015151515151515152
65	CRSPT vs. MS	26.189567799764497	3.493880774457326E-151	0.0015384615384615385	0.0018181818181818182
64	—DeepMARL-MR Prioritized (S)-model-5 vs. WINQ	23.87972265010945	4.9757581244479696E-126	0.0015625	0.0018181818181818182
63	—DeepMARL-MR Prioritized (S)-model-5 vs. MS	22.090724936161074	3.8811750949767075E-108	0.0015873015873015873	0.0018181818181818182
62	—DeepMARL-MR (S)-model-4 vs. WINQ	21.773687366600512	4.12106705278224E-105	0.0016129032258064516	0.0018181818181818182
61	CRSPT vs. —DeepMARL-AS (8 rules) (S)-model-1	21.07167560543091	1.4471874290747004E-98	0.0016393442622950820	0.0018181818181818182
60	—DeepMARL-MR (S)-model-4 vs. MS	19.98468965265214	7.48511854269219E-89	0.0016666666666666668	0.0018181818181818182
59	SPT vs. WINQ	19.73558870514025	1.066968205209761E-86	0.0016949152542372883	0.0018181818181818182
58	CRSPT vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	19.452519446604125	2.7745761202980874E-84	0.0017241379310344483	0.0018181818181818182
57	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. WINQ	18.263628560752267	1.6121263764696985E-74	0.0017543859649122807	0.0018181818181818182
56	SPT vs. MS	17.946590991191876	5.102802922815227E-72	0.0017857142857142859	0.0018181818181818182
55	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) (S)-model-1	16.972832741827492	1.3048306912228623E-64	0.0018181818181818182	0.0018181818181818182
54	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. MS	16.474630846803894	5.582829897620006E-61	0.001851851851851852	0.002173913043478261
53	CRSPT vs. LWRK	16.259498210316366	1.9129822755784764E-59	0.0018867924528301887	0.002173913043478261
52	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	15.353676583000704	3.3466951445833262E-53	0.0019230769230769232	0.002173913043478261
51	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) (S)-model-1	14.866797458318556	5.414200488257255E-50	0.0019607843137254 9	0.002173913043478261
50	CRSPT vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	14.77621529558693	2.085609981452213E-49	0.002	0.002173913043478261
49	—DeepMARL-AS (Beefy)-model-7 vs. WINQ	14.028912453051483	1.0373008508511168E-44	0.0020408163265306124	0.002173913043478261
48	CRSPT vs. —DeepMARL-AS (Beefy)-model-7	13.949653060661385	3.1613947775814485E-44	0.0020833333333333333	0.002173913043478261
47	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	13.247641299491768	4.6560796370062935E-40	0.002127659574468085	0.002173913043478261
46	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. WINQ	13.202350218125941	8.503983897021281E-40	0.002173913043478261	0.002173913043478261
45	SPT vs. —DeepMARL-AS (8 rules) (S)-model-1	12.828698796858296	1.1325019874700095E-37	0.0022222222222222222	0.0022222222222222222
44	—DeepMARL-AS (Beefy)-model-7 vs. MS	12.23991473910311	1.9023236122627718E-34	0.002272727272727273	0.00256410256410256 4
43	—DeepMARL-MR Prioritized (S)-model-5 vs. LWRK	12.160655346712945	5.035311126415846E-34	0.002325581395348837	0.00256410256410256 4
42	LWRK vs. WINQ	11.719067303396503	1.0178644267536058E-31	0.002380952380952381	0.00256410256410256 4
41	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. MS	11.413352504177567	3.586350655302524E-30	0.0024390243902439024	0.00256410256410256 4
40	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) (S)-model-1	11.35673865247031	6.866023872019803E-30	0.0025	0.00256410256410256 4
39	SPT vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	11.209542638031506	3.6608007587407133E-29	0.002564102564102564	0.00256410256410256 4
38	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	10.677372431983507	1.2989630871013547E-26	0.0026315789473684 21	0.00270270270270270 3
37	—DeepMARL-MR (S)-model-4 vs. LWRK	10.054620063204009	8.765890702799013E-24	0.0027027027027027 03	0.00270270270270270 3
36	LWRK vs. MS	9.930069589448129	3.080404863646444E-23	0.002777777777777778	0.002777777777777778
35	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (Beefy)-model-7	9.850810197057964	6.799363611102757E-23	0.002857142857142857	0.0029411764705882353
34	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	9.737582493643522	2.0845432002701058E-22	0.0029411764705882353	0.0029411764705882353
33	CRSPT vs. —DeepMARL-AS (8 rules) (GDT)-model-0	9.714936952960603	2.6041013258025245E-22	0.0030303030303030303	0.0032258064516129032
32	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	8.571337148474571	1.0229029222807898E-17	0.003125	0.0032258064516129032
31	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. WINQ	8.526046067108744	1.51434974485467E-17	0.0032258064516129032	0.0032258064516129032
30	CRSPT vs. SPT	8.24297680857262	1.679781409886975 8E-16	0.0033333333333333335	0.0033333333333333335
29	SPT vs. LWRK	8.016521401743747	1.0878197848639693E-15	0.003448275862068966	0.003448275862068966
28	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (Beefy)-model-7	7.744774913549029	9.575131268231217E-15	0.0035714285714285718	0.0035714285714285718
27	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) (S)-model-1	7.122022544769528	1.0635465389391118E-12	0.0037037037037037037	0.0037037037037037037
26	—DeepMARL-AS (8 rules) (S)-model-1 vs. WINQ	6.906889908281955	4.95394079040604E-12	0.0038461538461538464	0.004
25	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. MS	6.737048353160372	1.6163640399222215E-11	0.004	0.004
24	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. LWRK	6.544561257355763	5.967030269282831E-11	0.0041666666666666667	0.0041666666666666667
23	SPT vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	6.53323848701431	6.436253076989309E-11	0.004347826086956522	0.004347826086956522
22	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. —DeepMARL-AS (8 rules) (S)-model-1	6.295460309843985	3.0648897808150574E-10	0.0045454545454545546	0.0045454545454545546
21	CRSPT vs. —DeepMARL-MR (S)-model-4	6.204878147112357	5.473929939087479E-10	0.004761904761904762	0.004761904761904762
20	SPT vs. —DeepMARL-AS (Beefy)-model-7	5.7066762520887675	1.1520360652344736E-8	0.005	0.005
19	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-AS (8 rules) (GDT)-model-0	5.616094089357181	1.9532232889899644E-8	0.005263157894736842	0.005263157894736842
18	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	5.502866385942738	3.736656126280505E-8	0.005555555555555556	0.005555555555555556
17	—DeepMARL-AS (8 rules) (S)-model-1 vs. MS	5.1178921943335824	3.0896911055583667E-7	0.0058823529411764705	0.0058823529411764705
16	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	5.061278342626325	4.164547093898287E-7	0.00625	0.00625
15	LWRK vs. —DeepMARL-AS (8 rules) (S)-model-1	4.8121773951145475	1.4929474357434183E-6	0.006666666666666667	0.006666666666666667
14	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	4.676304151017196	2.9209129269717458E-6	0.0071428571428571435	0.0071428571428571435
13	—DeepMARL-AS (8 rules) (GDT)-model-0 vs. —DeepMARL-AS (Beefy)-model-7	4.234716107700783	2.2884062091512816E-5	0.007692307692307693	0.007692307692307693
12	—DeepMARL-MR Prioritized (S)-model-5 vs. SPT	4.144133944969196	3.41100338379446E-5	0.008333333333333333	0.008333333333333333
11	CRSPT vs. —DeepMARL-MR Prioritized (S)-model-5	4.098842863603422	4.1522080403999844E-5	0.009090909090909092	0.009090909090909092
10	—DeepMARL-MR (S)-model-4 vs. —DeepMARL-AS (8 rules) (GDT)-model-0	3.5100588058482463	4.4800760165300123E-4	0.01	0.01
9	LWRK vs. —DeepMARL-AS (8 rules) Prioritized (S)-model-3	3.193021236287758	0.001407925919310537	0.011111111111111112	0.011111111111111112
8	—DeepMARL-AS (Beefy)-model-7 vs. LWRK	2.30984514965498	0.020896728657797497	0.0125	0.0125
7	—DeepMARL-MR Prioritized (S)-model-5 vs. —DeepMARL-MR (S)-model-4	2.106035283508935	0.035201287589215	0.014285714285714287	0.014285714285714287
6	—DeepMARL-MR (S)-model-4 vs. SPT	2.038098661460262	0.04154006700988629	0.016666666666666666	0.016666666666666666
5	MS vs. WINQ	1.7889977139483728	0.07361518372016992	0.02	0.02
4	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs. —DeepMARL-AS (8 rules) (S)-model-1	1.6191561588267895	0.10541366817859567	0.025	0.025
3	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs. LWRK	1.483282914729438	0.13799925973447852	0.03333333333333333	0.03333333333333333
2	SPT vs. —DeepMARL-AS (8 rules) (GDT)-model-0	1.4719601443879842	0.1410316405207138	0.05	0.05
1	—DeepMARL-AS (Beefy)-model-7 vs. —DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	0.8265622349255423	0.4084852298799626	0.1	0.1

Table 7: Adjusted p-values

i	hypothesis	unadjusted p	p_{Neme}	p_{Holm}	p_{Shaf}	p_{Berg}
1	CRSPT vs .WINQ	2.962603460225532E-172	1.955318283748851E-170	1.955318283748851E-170	1.955318283748851E-170	0.0
2	CRSPT vs .MS	3.493880774457326E-151	2.305961311141835E-149	2.271022503397262E-149	1.921634425915293E-149	0.0
3	—DeepMARL-MR Prioritized (S)-model-5 vs .WINQ	4.9757581244479696E-126	3.28400036213566E-124	3.1844851996467006E-124	2.7366669684463833E-124	0.0
4	—DeepMARL-MR Prioritized (S)-model-5 vs .MS	3.8811750949767075E-108	2.561575562684627E-106	2.445140309835326E-106	2.1346463022371891E-106	0.0
5	—DeepMARL-MR (S)-model-4 vs .WINQ	4.121067052782224E-105	2.719904254836268E-103	2.555061572724979E-103	2.266586879030223E-103	0.0
6	CRSPT vs .—DeepMARL-AS (8 rules) (S)-model-1	1.4471874290747004E-98	9.551437031893023E-97	8.82784331735567 2E-97	7.959530859910853E-97	0.0
7	—DeepMARL-MR (S)-model-4 vs .MS	7.48511854269219E-89	4.940178238176846E-87	4.491071256153 15E-87	4.116815198480705E-87	0.0
8	SPT vs .WINQ	1.066968205209761E-86	7.041990154384423E-85	6.29511241073759E-85	5.868325128653685E-85	0.0
9	CRSPT vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	2.7745761202980874E-84	1.8312202393967376E-82	1.609254149772890 8E-82	1.526016866163948E-82	0.0
10	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .WINQ	1.6121263764696985E-74	1.0640034084700011E-72	9.189120345877281E-73	8.866695070583342E-73	0.0
11	SPT vs .MS	5.102802922815227E-72	3.36784992905805E-70	2.857569636776527E-70	2.806541607548375E-70	0.0
12	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-AS (8 rules) (S)-model-1	1.3048306912228623E-64	8.611882562070891E-63	7.176568801725742E-63	7.176568801725742E-63	0.0
13	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .MS	5.582829897620006E-61	3.684667732492204E-59	3.014728144714803 2E-59	2.568101752905203E-59	0.0
14	CRSPT vs .LWRK	1.9129822757584764E-59	1.2625683020005944E-57	1.013880606151992 5E-57	8.799718468488992E-58	0.0
15	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	3.3466951445833262E-53	2.2088187954249953E-51	1.740281475183329 7E-51	1.53947976650833E-51	0.0
16	—DeepMARL-MR (S)-model-4 vs .—DeepMARL-AS (8 rules) (S)-model-1	5.414200488257255E-50	3.5733723222497883E-48	2.761242249011 2E-48	2.490532224598337 6E-48	0.0
17	CRSPT vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	2.085609981452213E-49	1.3765025877584607E-47	1.042804990726106 5E-47	9.59380591468018E-48	0.0
18	—DeepMARL-AS (Beefy)-model-7 vs .WINQ	1.0373008508511168E-44	6.846185615617371E-43	5.082774169170472E-43	4.771583913915375E-43	0.0
19	CRSPT vs .—DeepMARL-AS (Beefy)-model-7	3.1613947775814485E-44	2.086520553203756E-42	1.517469493239095 3E-42	1.454241597687466 3E-42	0.0
20	—DeepMARL-MR (S)-model-4 vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	4.6560796370062935E-40	3.0730125604241536E-38	2.188357429392958E-38	2.141796633022895E-38	0.0
21	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs .WINQ	8.503983897021281E-40	5.612629372034045E-38	3.911832592629789 5E-38	3.911832592629789 5E-38	0.0
22	SPT vs .—DeepMARL-AS (8 rules) (S)-model-1	1.1325019874700095E-37	7.474513117302062E-36	5.096258943615043E-36	5.096258943615043E-36	0.0
23	—DeepMARL-AS (Beefy)-model-7 vs .MS	1.9023236122627718E-34	1.2555335840934294E-32	8.370223893956196E-33	7.419062087824811E-33	0.0
24	—DeepMARL-MR Prioritized (S)-model-5 vs .LWRK	5.035311126415846E-34	3.3233053434344587E-32	2.165183784358814E-32	1.963771339302180 2E-32	0.0
25	LWRK vs .WINQ	1.0178644267536058E-31	6.717905216573799E-30	4.275030592365144E-30	3.969671264339063E-30	0.0
26	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs .MS	3.586350655302524E-30	2.366991432499666E-28	1.470403768674035E-28	1.398676755567984 5E-28	0.0
27	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .—DeepMARL-AS (8 rules) (S)-model-1	6.866023872019803E-30	4.531575755533069 5E-28	2.746409548807921E-28	2.677749310087723E-28	0.0
28	SPT vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	3.6608007587407133E-29	2.416128500768871E-27	1.427712295908878 2E-27	1.427712295908878 2E-27	0.0
29	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	1.2989630871013547E-26	8.573156374868941E-25	4.936059730985148E-25	4.806163422275012E-25	0.0
30	—DeepMARL-MR (S)-model-4 vs .LWRK	8.765890702799013E-24	5.785487863847349E-22	3.243379560035635E-22	3.243379560035635E-22	0.0
31	LWRK vs .MS	3.080404863646444E-23	2.033067210006653E-21	1.108945750912719 7E-21	1.108945750912719 7E-21	0.0
32	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-AS (Beefy)-model-7	6.7993636111027 57E-23	4.48757998332782E-21	2.379777263885965E-21	2.311783627749374E-21	0.0
33	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	2.0845432002701058E-22	1.3757985121782698E-20	7.08744688091836E-21	7.08744688091836E-21	0.0
34	CRSPT vs .—DeepMARL-AS (8 rules) (GDT)-model-0	2.6041013258025245E-22	1.7187068750296663E-20	8.593534375148331E-21	8.072714109987825E-21	0.0
35	—DeepMARL-MR (S)-model-4 vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	1.0229029222807898E-17	6.751159287053212E-16	3.273289351298527 5E-16	3.170999059070448 7E-16	0.0
36	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs .WINQ	1.514349744854 67E-17	9.994708316040 82E-16	4.694484209049477E-16	4.694484209049477E-16	0.0
37	CRSPT vs .SPT	1.6797814098869758E-16	1.108655730525404E-14	5.039344229660927 5E-15	5.039344229660927 5E-15	0.0
38	SPT vs .LWRK	1.0878197848639693E-15	7.179610580102197E-14	3.154677376105510 7E-14	3.154677376105510 7E-14	0.0
39	—DeepMARL-MR (S)-model-4 vs .—DeepMARL-AS (Beefy)-model-7	9.575131268231217E-15	6.319586637032603E-13	2.681036755104741E-13	2.681036755104741E-13	0.0
40	—DeepMARL-AS (Beefy)-model-7 vs .—DeepMARL-AS (8 rules) (S)-model-1	1.0635465389391118E-12	7.019407156998137E-11	2.871575655135601 7E-11	2.871575655135601 7E-11	0.0
41	—DeepMARL-AS (8 rules) (S)-model-1 vs .WINQ	4.95394079040604E-12	3.2696009216679865E-10	1.288024605505570 5E-10	1.23848519760151E-10	0.0
42	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs .MS	1.6163640399222215E-11	1.0668002663486662E-9	4.040910099805553 6E-10	4.040910099805553 6E-10	0.0
43	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .LWRK	5.967030269282831E-11	3.938239977726669E-9	1.432087264627879 4E-9	1.432087264627879 4E-9	0.0
44	SPT vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	6.436253076989309E-11	4.247927030812944 5E-9	1.480338207707541E-9	1.480338207707541E-9	0.0
45	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs .—DeepMARL-AS (8 rules) (S)-model-1	3.0648897808150574E-10	2.022827255337937 8E-8	6.742757517793126 5E-9	6.742757517793126 5E-9	0.0
46	CRSPT vs .—DeepMARL-MR (S)-model-4	5.473929939087479E-10	3.612793759797736E-8	1.149525287208370 6E-8	1.149525287208370 6E-8	0.0
47	SPT vs .—DeepMARL-AS (Beefy)-model-7	1.1520360652344736E-8	7.603438030547525E-7	2.304072130468947 2E-7	2.304072130468947 2E-7	0.0
48	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-AS (8 rules) (GDT)-model-0	1.9532232889899644E-8	1.2891273707333764E-6	3.711124249080932E-7	3.711124249080932E-7	0.0
49	—DeepMARL-AS (Beefy)-model-7 vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	3.736656126280505E-8	2.4661930433451334E-6	6.725981027304909E-7	6.725981027304909E-7	0.0
50	—DeepMARL-AS (8 rules) (S)-model-1 vs .MS	3.0896911055583667E-7	2.039196129668522E-5	5.252474879449223E-6	5.252474879449223E-6	0.0
51	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	4.164547093898287E-7	2.7486010819728693E-5	6.663275350237259E-6	6.663275350237259E-6	0.0
52	LWRK vs .—DeepMARL-AS (8 rules) (S)-model-1	1.4929474357434183E-6	9.85345307590656E-5	2.239421153615127 6E-5	2.239421153615127 6E-5	0.0
53	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	2.9209129269717458E-6	1.9278025318013523E-4	4.089278097604444E-5	4.089278097604444E-5	0.0
54	—DeepMARL-AS (8 rules) (GDT)-model-0 vs .—DeepMARL-AS (Beefy)-model-7	2.2884062091512816E-5	0.0015103480980398459	2.974928071896666E-4	2.974928071896666E-4	0.0
55	—DeepMARL-MR Prioritized (S)-model-5 vs .SPT	3.41100338379446E-5	0.0022512622333043437	4.093204060553352E-4	4.093204060553352E-4	0.0
56	CRSPT vs .—DeepMARL-MR Prioritized (S)-model-5	4.1522080403999844E-5	0.0027404573066639897	4.567428844439983E-4	4.567428844439983E-4	0.0
57	—DeepMARL-MR (S)-model-4 vs .—DeepMARL-AS (8 rules) (GDT)-model-0	4.4800760165300123E-4	0.029568501709098083	0.004480076016530012	0.004480076016530012	0.0
58	LWRK vs .—DeepMARL-AS (8 rules) Prioritized (S)-model-3	0.001407925919310537	0.09292311067449543	0.012671333273794832	0.012671333273794832	0.0
59	—DeepMARL-AS (Beefy)-model-7 vs .LWRK	0.020896728657797497	1.3791840914146347	0.16717382926237997	0.16717382926237997	0.0
60	—DeepMARL-MR Prioritized (S)-model-5 vs .—DeepMARL-MR (S)-model-4	0.035201287589215	2.32328490888819	0.246409013124505	0.246409013124505	0.0
61	—DeepMARL-MR (S)-model-4 vs .SPT	0.04154006700988629	2.741644422652495	0.24924040205931774	0.24924040205931774	0.0
62	MS vs .WINQ	0.07361518372016992	4.858602125531215	0.368075918600849 65	0.368075918600849 65	0.0
63	—DeepMARL-AS (8 rules) Prioritized (S)-model-3 vs .—DeepMARL-AS (8 rules) (S)-model-1	0.10541366817859567	6.957302099787314	0.4216546727143827	0.4216546727143827	0.0
64	—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2 vs .LWRK	0.13799925973447852	9.107951142475581	0.4216546727143827	0.4216546727143827	0.0
65	SPT vs .—DeepMARL-AS (8 rules) (GDT)-model-0	0.1410316405207138	9.30808827436711	0.4216546727143827	0.4216546727143827	0.0
66	—DeepMARL-AS (Beefy)-model-7 vs .—DeepMARL-AS (8 rules) Prioritized (GDT)-model-2	0.4084852298799626	26.960025172077533	0.4216546727143827	0.4216546727143827	0.0