

Table 1: DL + non-DL results with different weight treatments. Cells show medians of 10 runs. **Dark blue** shows top rank (note: for pf, *less is better*), **light blue** shows rank two; white shows lowest rank (worst performance). Rankings were calculated via Scott-Knott. The train and test versions used here are the same as the DODGE paper.

Dataset	Treatment	AUC	f1	recall	pf
cloudstac-1 day	none	0.76	0.714	1.0	0.191
	weighted	0.755	0.709	1.0	0.2
	all	0.76	0.728	1.0	0.287
cloudstac-14 days	none	0.643	0.195	0.667	0.0
	weighted	0.643	0.195	0.667	0.0
	all	0.568	0.157	0.92	0.011
cloudstac-180 days	none	0.791	0.211	1.0	0.0
	weighted	0.791	0.286	1.0	0.0
	all	0.748	0.258	1.0	0.016
cloudstac-30 days	none	0.647	0.222	1.0	0.0
	weighted	0.647	0.216	0.5	0.0
	all	0.609	0.286	0.917	0.036
cloudstac-365 days	none	0.624	0.333	1.0	0.003
	weighted	0.624	0.333	1.0	0.003
	all	0.624	0.333	1.0	0.003
cloudstac-7 days	none	0.611	0.486	0.968	0.007
	weighted	0.611	0.486	0.968	0.007
	all	0.658	0.507	0.968	0.19
cloudstac-90 days	none	0.679	0.267	0.923	0.019
	weighted	0.679	0.286	0.923	0.011
	all	0.631	0.286	1.0	0.056
cocoo-14 days	none	0.806	0.2	0.929	0.002
	weighted	0.806	0.143	0.929	0.002
	all	0.887	0.15	1.0	0.01
cocoo-180 days	none	0.809	0.143	0.9	0.0
	weighted	0.809	0.143	0.9	0.0
	all	0.856	0.211	1.0	0.014
cocoo-30 days	none	0.835	0.126	1.0	0.0
	weighted	0.835	0.126	1.0	0.0
	all	0.88	0.129	1.0	0.014
cocoo-365 days	none	0.721	0.17	0.95	0.0
	weighted	0.721	0.195	0.95	0.0
	all	0.813	0.179	0.963	0.027
cocoo-7 days	none	0.829	0.294	1.0	0.006
	weighted	0.829	0.34	1.0	0.006
	all	0.845	0.348	1.0	0.024
cocoo-90 days	none	0.734	0.308	1.0	0.331
	weighted	0.734	0.276	1.0	0.362
	all	0.827	0.4	1.0	0.378
deeplearnin-1 day	none	0.755	0.842	1.0	0.003
	weighted	0.746	0.851	1.0	0.003
	all	0.725	0.841	1.0	0.024
deeplearnin-14 days	none	0.524	0.118	1.0	0.002
	weighted	0.524	0.118	1.0	0.002
	all	0.524	0.118	1.0	0.002

Table 2: Non-DL results with different weight treatments. Cells show medians of 10 runs. **Dark blue** shows top rank (note: for pf, *less is better*), **light blue** shows rank two; white shows lowest rank (worst performance). Rankings were calculated via Scott-Knott. The train and test versions used here are the same as the DODGE paper.

Dataset	Treatment	AUC	f1	recall	pf
cloudstac-1 day	none	0.76	0.714	0.722	0.191
	weighted	0.755	0.709	0.715	0.2
	all	0.76	0.728	0.981	0.287
cloudstac-14 days	none	0.643	0.195	0.6	0.0
	weighted	0.643	0.195	0.6	0.0
	all	0.568	0.157	0.92	0.011
cloudstac-180 days	none	0.791	0.211	0.867	0.0
	weighted	0.791	0.286	0.867	0.0
	all	0.748	0.258	1.0	0.016
cloudstac-30 days	none	0.647	0.222	0.5	0.0
	weighted	0.647	0.216	0.5	0.0
	all	0.609	0.286	0.917	0.036
cloudstac-365 days	none	0.624	0.333	0.25	0.003
	weighted	0.624	0.333	0.25	0.003
	all	0.624	0.333	0.5	0.003
cloudstac-7 days	none	0.611	0.486	0.771	0.007
	weighted	0.611	0.486	0.771	0.007
	all	0.658	0.507	0.982	0.19
cloudstac-90 days	none	0.679	0.267	0.69	0.019
	weighted	0.679	0.286	0.69	0.011
	all	0.631	0.286	0.931	0.056
cocoo-1 day	none	0.851	0.348	0.923	0.004
	weighted	0.851	0.34	0.923	0.016
	all	0.884	0.406	1.0	0.058
cocoo-14 days	none	0.806	0.2	0.8	0.002
	weighted	0.806	0.143	0.8	0.002
	all	0.887	0.15	1.0	0.01
cocoo-180 days	none	0.809	0.143	0.929	0.0
	weighted	0.809	0.143	0.929	0.0
	all	0.856	0.211	1.0	0.006
cocoo-30 days	none	0.835	0.126	0.889	0.0
	weighted	0.835	0.126	0.889	0.0
	all	0.88	0.129	1.0	0.014
cocoo-365 days	none	0.721	0.17	0.75	0.0
	weighted	0.721	0.195	0.75	0.0
	all	0.813	0.179	1.0	0.014
cocoo-7 days	none	0.829	0.294	0.889	0.0
	weighted	0.829	0.34	0.889	0.0
	all	0.845	0.348	0.963	0.027
cocoo-90 days	none	0.734	0.308	0.722	0.006
	weighted	0.734	0.276	0.722	0.006
	all	0.827	0.4	0.944	0.024
cocoo-14 days	none	0.755	0.842	0.918	0.331
	weighted	0.755	0.842	0.918	0.331
	all	0.755	0.842	0.918	0.331