

Supplementary Table 3. Cleanlab score association with confirmed label errors across nine datasets.

Dataset	Module	N (Total)	N (Error)	Error Rate	Mann-Whitney p-value	AUC
EDC	Label Issue	133	68	51.13%	0.55	0.53
	Outlier	82	25	30.49%	0.02	0.67
MuReD	Label Issue	479	231	48.23%	0.01	0.57
	Outlier	12	2	16.67%	0.91	0.55
ODIR	Label Issue	1301	512	39.35%	0.01	0.54
	Outlier	18	7	38.89%	0.72	0.56
OCT2017	Label Issue	355	133	37.46%	0.06	0.56
	Outlier	30	11	36.67%	0.12	0.67
OCTID	Label Issue	7	4	57.14%	0.06	1
	Outlier	/	/	/	/	/
OCTDL	Label Issue	97	33	34.02%	0.53	0.54
	Outlier	16	4	25.00%	0.95	0.52
OpenUWF	Label Issue	103	41	39.81%	0.07	0.61
	Outlier	35	5	14.29%	0.14	0.71
tumorUWF	Label Issue	44	1	2.27%	0.95	0.53
	Outlier	102	1	0.98%	0.37	0.82
TsukazakiUWF	Label Issue	1248	714	57.21%	<0.01	0.57
	Outlier	599	189	31.55%	0.24	0.53