

S2 Table The paired t-test results between models built with the two different binaries metrics.

Classifier	Binary metric	Accuracy	Balanced Accuracy	Cohen's Kappa	F ₁ -score	MCC	DOR
PLS-DA	Accuracy	0.830	0.813	0.616	0.743	0.617	20.150
SVM linear	Accuracy	0.764	0.741	0.471	0.648	0.472	8.662
SVM poly	Accuracy	0.792	0.746	0.508	0.656	0.510	11.308
SVM radial	Accuracy	0.792	0.801	0.558	0.718	0.570	16.333
RF	Accuracy	0.830	0.735	0.547	0.640	0.614	-
PLS-DA	ROC	0.792	0.777	0.538	0.694	0.540	12.607
SVM linear	ROC	0.745	0.704	0.411	0.597	0.411	6.484
SVM poly	ROC	0.764	0.725	0.455	0.627	0.455	8.077
SVM radial	ROC	0.774	0.740	0.480	0.647	0.480	9.167
RF	ROC	0.811	0.706	0.487	0.583	0.568	-
<i>p</i> -value		0.0029	0.0051	0.0002	0.0002	0.0010	0.0344

The *p*-value comes from calculating a t-paired test ($H_0: \mu=0$, $H_a: \mu \neq 0$) to compare confusion matrix metrics from a model built with Accuracy or ROC. The abbreviations signify DOR (Diagnostic odd ratio) and MCC (Matthews correlation coefficient).