

CORRECTION

# Correction: Novel comparison of evaluation metrics for gene ontology classifiers reveals drastic performance differences

Ilya Plyusnin, Liisa Holm, Petri Törönen

There are several errors in [Table 1](#). The values for the column rec in rows ic SimGIC2, ic2 SimGIC2, AJacc E, and ic2 S<sub>min1</sub> are incorrect. Please see the correct [Table 1](#) here.

**Table 1. Summary of results for best performing and widely-used metrics.** Here we show RC (Rank Correlation) and FP (False Positive) results for the best performing methods. We also show some results for some widely-used metrics. Good metrics should have a high RC score and low FP scores. Rec column shows our selected recommendations (See text for details). The five best results in each column are shown in bold. The five weakest results in each column are shown with underlined italics. Metrics that fail a given test are highlighted in red (see text for details). Note how methods in lower block show consistent weak performance either in RC or FP tests.

Top performing metrics								
Metric	Rank Correlation Results			False Positive Sets			rec	weakness
	UniProt	CAFA	Mouse	UniProt	CAFA	Mouse		
TC AUCROC	0.959	0.951	<i>0.920</i>	<b>0.023</b>	<b>0.010</b>	<b>0.000</b>	(*)	RC in mouse data
TC AUCPR	<b>0.984</b>	<b>0.982</b>	0.971	0.144	0.156	<i>0.312</i>	(*)	FPS in mouse data
ic SimGIC	0.960	0.963	0.969	0.168	0.164	0.108		
ic SimGIC2	0.970	0.969	0.965	0.166	0.133	<b>0.056</b>	*	
ic2 SimGIC2	<b>0.979</b>	0.978	0.974	0.190	0.156	0.065	(*)	
Resnik E	0.966	0.959	0.945	<b>0.034</b>	<b>0.064</b>	0.065	(*)	RC in mouse data
Lin E	0.939	<b>0.983</b>	<b>0.982</b>	0.112	0.118	0.080	(*)	
Lin F	<i>0.856</i>	0.979	<b>0.978</b>	<b>0.096</b>	<b>0.100</b>	<b>0.058</b>		very weak RC in Uniprot
Ajacc E	<i>0.886</i>	0.960	0.959	0.097	0.127	0.073		very weak RC in Uniprot
ic2 S <sub>min1</sub>	<b>0.986</b>	<b>0.985</b>	<b>0.983</b>	0.247	0.221	0.124	(*)	Slightly weak in FPS
Previously used metrics with weaker performance								
F <sub>max</sub>	<b>0.983</b>	<b>0.982</b>	<b>0.981</b>	0.367	0.318	0.229		weak in FPS tests
US AUCPR	<b>0.985</b>	<b>0.983</b>	<b>0.977</b>	<i>0.453</i>	<i>0.388</i>	<i>0.292</i>		weak in FPS tests
US AUCROC	0.945	<i>0.932</i>	<i>0.901</i>	<i>1.000</i>	<i>1.000</i>	<i>0.878</i>		worst metrics in FPS tests
GC AUCROC	0.947	0.937	0.921	<i>1.000</i>	<i>1.000</i>	<i>0.879</i>		worst metrics in FPS tests
Resnik A	0.922	<i>0.808</i>	<i>0.772</i>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>		weak in RC tests
Resnik D	<i>0.892</i>	<i>0.811</i>	<i>0.801</i>	<i>0.428</i>	<i>0.333</i>	<i>0.339</i>		weak in all tests
Lin A	<i>0.758</i>	<i>0.840</i>	<i>0.880</i>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>		worst metrics in RC test
Lin D	<i>0.806</i>	<i>0.926</i>	0.970	<i>0.466</i>	<i>0.425</i>	<i>0.364</i>		weak in all tests



**OPEN ACCESS**

**Citation:** Plyusnin I, Holm L, Törönen P (2022) Correction: Novel comparison of evaluation metrics for gene ontology classifiers reveals drastic performance differences. PLoS Comput Biol 18(6): e1010249. <https://doi.org/10.1371/journal.pcbi.1010249>

**Published:** June 9, 2022

**Copyright:** © 2022 Plyusnin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.1371/journal.pcbi.1010249.t001>

**Reference**

1. Plyusnin I, Holm L, Törönen P (2019) Novel comparison of evaluation metrics for gene ontology classifiers reveals drastic performance differences. PLoS Comput Biol 15(11): e1007419. <https://doi.org/10.1371/journal.pcbi.1007419> PMID: 31682632