

Table 12 and Table13 show the results of our statistical analysis of the four RQs in Section 5. In Table 12, the first two rows are the results of RQ1, the 3rd row is the results of RQ2, rows 4 to 11 are the results of RQ3. The Table 13 represents the results of RQ4. Specifically, each row shows the median values of the p-value and delta between DeMuVGN and baselines for all datasets on each measure.

Table 12. Statistical comparison of DeMuVGN with baselines in the within-project context.

Approaches	Measures	AUC	Recall	Brier	PF	F1
BiGGNN vs CDG-BiGGNN	p-value	3.181e-02(L)	1.370e-02(L)	6.134e-06(L)	8.453e-02	4.324e-07(L)
	delta	0.962(L)	0.943(L)	0.915(L)	0.580(L)	1.000(L)
BiGGNN vs DDG-BiGGNN	p-value	1.358e-02(L)	1.211e-04(L)	1.026e-07(L)	2.607e-01	2.950e-06(L)
	delta	1.000(L)	0.960(L)	0.845(L)	0.692(L)	1.000(L)
BiGGNN vs BiGGNN	p-value	4.132e-02(L)	2.384e-02(L)	2.136e-04(L)	2.118e-01	6.084e-04(L)
	delta	0.915(L)	0.980(L)	0.922(L)	0.345(L)	1.000(L)
BiGGNN vs DBN	p-value	6.076e-08(L)	1.656e-07(L)	1.147e-02(L)	3.662e-03(L)	1.817e-07(L)
	delta	1.000(L)	1.000(L)	0.815(L)	0.525(L)	1.000(L)
BiGGNN vs CGCN	p-value	4.906e-05(L)	7.665e-07(L)	1.814e-01	6.976e-03(L)	2.240e-07(L)
	delta	0.938(L)	0.958(L)	0.742(L)	0.431(L)	0.988(L)
BiGGNN vs DP-CNN	p-value	6.076e-08(L)	1.656e-07(L)	8.512e-08(L)	3.447e-05(L)	1.466e-07(L)
	delta	1.000(L)	1.000(L)	0.807(L)	0.655(L)	1.000(L)
BiGGNN vs LR+	p-value	6.076e-08(L)	1.800e-07(L)	2.552e-07(L)	3.280e-06(L)	1.466e-07(L)
	delta	1.000(L)	0.985(L)	0.740(L)	0.715(L)	1.000(L)
BiGGNN vs NB+	p-value	6.076e-08(L)	1.656e-07(L)	4.158e-05(L)	3.219e-05(L)	1.720e-07(L)
	delta	1.000(L)	1.000(L)	0.770(L)	0.490(L)	1.000(L)
BiGGNN vs RF+	p-value	6.076e-08(L)	1.656e-07(L)	4.580e-03(L)	3.686e-02(L)	1.466e-07(L)
	delta	0.995(L)	1.000(L)	0.587(L)	0.411(L)	1.000(L)
BiGGNN vs SVM+	p-value	6.076e-08(L)	1.656e-07(L)	1.577e-07(L)	3.765e-05(L)	1.466e-07(L)
	delta	0.911(L)	0.999(L)	0.502(L)	0.490(L)	1.000(L)
BiGGNN vs XGBoost+	p-value	6.076e-08(L)	1.656e-07(L)	1.308e-03(L)	1.996e-03(L)	1.466e-07(L)
	delta	1.000(L)	1.000(L)	0.828(L)	0.561(L)	1.000(L)

Table 13. Statistical comparison of DeMuVGN with baselines in the cross-project context.

Approaches	Measures	Recall	pf	Brier	AUC	F1
DeMuVGN vs CGCN	p-value	1.399E-03(L)	7.147E-04(L)	3.828E-07(L)	3.090E-08(L)	2.508E-05(L)
	delta	0.633(L)	0.571(L)	0.714(L)	0.837(L)	1.000(L)
DeMuVGN vs DSSDPP	p-value	1.107E-07(L)	3.895E-05(L)	2.961E-08(L)	3.090E-08(L)	3.113E-08(L)
	delta	0.653(L)	0.653(L)	0.429(L)	0.755(L)	1.000(L)
DeMuVGN vs MNB	p-value	3.209E-04(L)	2.676E-07(L)	3.454E-08(L)	3.100E-08(L)	3.060E-06(L)
	delta	0.694(L)	0.592(L)	0.714(L)	0.837(L)	1.000(L)
DeMuVGN vs TCB	p-value	2.646E-04(L)	2.169E-04(L)	2.973E-08(L)	3.090E-08(L)	1.105E-07(L)
	delta	0.714(L)	0.531(L)	0.694(L)	0.755(L)	1.000(L)
DeMuVGN vs VCB-SVM	p-value	2.878E-06(L)	6.333E-08(L)	2.988E-08(L)	3.100E-08(L)	3.113E-08(L)
	delta	0.918(L)	0.735(L)	0.755(L)	0.918(L)	1.000(L)