

December 9 - December 16 Summary

[Spreadsheet link](#) | [Raise an issue](#)

- Improved results a bit by using wider neural nets, but more neurons.

		Standard deep learning tricks					Our method					MSR				Lin Tan			
		W	P	R	F	Runtime	Model	P	R	F	Runtime	P	R	F	Runtime	P	R	F	Runtime
PROMISE	ant 1.5-1.6		52.9 (1.2)	59.8 (3.2)	55.6 (0.3)	3.4 (0.2)	X11012310	51.9 (3.7)	72.8 (10.9)	59.9 (1.9)	20.6 (0.4)	33	80	47		44.8	51.1	47.7	
	ant 1.6-1.7		47.7 (2.3)	63.3 (6.6)	54.2 (0.8)	5.8 (0.6)	X12013610W	43.2 (1.5)	78.3 (1.8)	55.7 (0.4)	3.4 (0.3)	21	98	35		41.8	77.1	54.2	
	camel 1.2-1.4		24.5 (1.4)	64.1 (5.6)	35.3 (1.2)	9.7 (0.7)	X11002310/W	25.6 (1.3)	55.2 (4.8)	34.6 (1.4)	8.6 (0.8)	20	82	32		24.8	75.2	37.3	
	camel 1.4-1.6	Yes	35.2 (1.6)	51.1 (2.9)	40.7 (1.0)	15.0 (0.5)	X11002310	28.6 (1.1)	48.9 (5.9)	35.9 (1.6)	31.5 (0.2)	28	68	40		28.3	63.7	39.1	
	ivy 1.4-2.0	Tried	26.3 (4.7)	37.5 (5.0)	30.9 (3.1)	13.3 (0.6)	X11012610	20.4 (2.7)	67.5 (10.0)	32.3 (3.9)	43.6 (0.8)	11	70	18		22.6	60	32.9	
	jedit 3.2-4.0	Tried	41.1 (0.4)	72.0 (4.0)	51.3 (1.7)	15.4 (0.6)	X11002310W	43.2 (0.6)	66.7 (2.6)	51.4 (1.4)	8.5 (0.4)	35	75	47		44.7	73.3	55.6	
	jedit 4.0-4.1		50.0 (1.7)	68.4 (3.8)	57.4 (1.1)	18.5 (0.5)	X11002310	47.8 (2.7)	64.6 (5.0)	55.2 (1.4)	14.8 (0.5)	33	93	49		46.1	67.1	54.6	
	log4j 1.0-1.1		63.2 (1.1)	73.0 (2.7)	68.3 (0.1)	15.6 (0.4)	X11002310	61.9 (1.5)	67.6 (2.7)	64.9 (2.8)	5.7 (0.5)	37	97	52	Hours to	49.1	73	58.7	
	lucene 2.0-2.2	Tried	68.6 (0.8)	55.6 (0.6)	60.4 (2.1)	32.9 (0.5)	X12013610	60.5 (1.1)	64.6 (9.7)	59.3 (6.9)	21.2 (0.1)	60	100	75	days	73.3	38.2	50.2	200 iterations, 15 seconds
	lucene 2.2-2.4		70.5 (1.3)	65.0 (2.0)	66.7 (1.5)	21.1 (0.4)	X11002301W	64.3 (2.1)	88.2 (2.9)	74.1 (1.1)	15.5 (0.3)	60	100	75		70.9	52.7	60.5	
	poi 1.5-2.5	Yes	83.4 (1.4)	80.6 (8.2)	82.4 (3.5)	14.8 (0.6)	X11002310	81.1 (1.9)	77.8 (2.0)	79.6 (1.2)	28.0 (0.7)	70	96	81		73.7	44.8	55.8	
	poi 2.5-3.0		78.7 (1.4)	70.1 (2.9)	74.5 (1.3)	25.0 (0.7)	X11002301W	67.8 (2.8)	92.5 (2.9)	78.3 (0.7)	13.2 (0.2)	65	97	78		75	75.8	75.4	
	synapse 1.0-1.1	Tried	47.2 (4.7)	35.0 (1.7)	38.2 (4.1)	43.9 (0.2)	X11013310	32.7 (5.1)	58.3 (11.7)	42.5 (1.2)	39.2 (0.3)	30	88	46		45.5	50	47.6	
	synapse 1.1-1.2		48.9 (4.4)	44.2 (4.6)	45.4 (1.9)	67.6 (0.6)	X12013610	48.1 (1.9)	57.0 (12.8)	57.1 (1.3)	51.4 (0.4)	37	93	51		51.1	55.8	53.3	
	xalan 2.4-2.5		63.4 (0.5)	37.7 (2.6)	46.9 (2.8)	168.5 (1.3)	X11012310/W	62.6 (0.7)	47.3 (3.6)	54.6 (1.8)	113.9 (0.3)	49	100	66		64.7	43.2	51.8	
	xerces 1.2-1.3		15.5 (1.4)	29.0 (14.5)	20.5 (3.2)	22.5 (0.5)	X11012310	16.1 (1.0)	75.4 (17.4)	31.3 (3.0)	43.9 (0.3)	23	28	26		16	46.4	23.8	

- We should instead compare (deep learning literature + our method) vs. (MSR + Lin Tan), i.e., standard deep learning stuff vs. SE literature

Ours	Theirs
59.9	47
55.7	35
34.6	37.3
40.7	40
32.3	32.9
51.4	55.6
57.4	49
68.3	52
59.3	75
74.1	75
82.4	81
78.3	78
42.5	47.6
57.1	51
54.6	66
31.3	26

- Overall Cohen d (with bias correction): 0.35 (medium effect)
- Overall Cohen d (without bias correction): 0.40 (medium effect)
- Individual *diff / pooled SD* values:

Ours	Theirs	Difference
59.9	47	12.9
55.7	35	20.7
34.6	37.3	-2.7
40.7	40	0.7
32.3	32.9	-0.6
51.4	55.6	-4.2
57.4	49	8.4
68.3	52	16.3
59.3	75	-15.7
74.1	75	-0.9
82.4	81	1.4
78.3	78	0.3
42.5	47.6	-5.1
57.1	51	6.1
54.6	66	-11.4
31.3	26	5.3

- Started running Cross-Project Defect Prediction (CPDP)

Data	Our method	MSR	Lin Tan	TCA+
ant 1.6 - camel 1.4	33.2 (2.4)	32	31.6	29.2
jEdit 4.1 - camel 1.4	32.4 (1.9)	31	69.3	33
camel 1.4 - ant 1.6	57.0 (2.3)	45	97.9	61.1
poi 3.0 - ant 1.6	57.4 (1.9)	39	47.8	59.8