Table XIV: Overall performance of KillBadCode and baselines in detecting code poisoning. F: FPR; P: Precision; R: Recall. BC BadCode; CP: CodePoisoner.

Code Poisoning	AC					SS					ONION					CodeDetector					KillBadCode				
	F(%)	R(%)	P (%)	F1(%)	Time	F (%)	R(%)	P(%)	F1(%)	Time	F(%)	R(%)	P(%)	F1(%)	Time	F(%)	R(%)	P(%)	F1(%)	Time	F(%)	R(%)	P(%)	F1(%)	Time
Defect Detection																									
BC (Fixed)	9.06	30.71	77.14	43.93	0h37m	16.30	11.02	57.88	18.38	0h36m	67.64	35.02	9.41	14.87	23h15m	0	0	0	0	8h15m	3.81	100	96.42	98.18	0h20m
BC (Mixed)	24.58	36.93	61.28	46.11	0h37m	12.13	15.68	56.16	24.32	0h36m	68.48	27.68	8.56	13.23	23h15m	0	0	0	0	7h53m	5.18	100	95.08	97.48	0h20m
BNC (Fixed)	27.51	28.57	51.37	36.68	0h37m	24.23	11.27	32.89	16.54	0h36m	62.31	13.92	6.04	8.55	23h15m	0	0	0	0	7h27m	3.03	100	95.02	97.43	0h20m
BNC (Grammar)	25.72	25.71	50.33	34.12	0h37m	8.49	44.57	84.61	58.36	0h36m	71.81	19.52	7.73	11.04	23h15m	0	0	0	0	7h32m	14.88	100	85.12	91.92	0h20m
CP (Variable)	43.96	14.27	20.48	17.02	0h37m	4.58	48.03	84.73	61.43	0h36m	75.73	29.24	9.58	14.49	23h15m	0	0	0	0	8h57m	23.43	100	77.56	87.36	0h20m
Average	26.17	27.24	42.05	35.57	0h37m	13.15	26.11	63.25	35.81	0h36m	69.19	25.08	8.26	12.44	23h15m	0	0	0	0	8h57m	10.07	100	89.84	94.47	0h20m
Clone Detection																									
BC (Fixed)	49.38	0	0	0	4h31m	1.53	2.25	57.21	4.34	4h27m	64.55	37.52	18.30	24.56	17h21m	34.49	100	65.51	79.15	4h40m	2.50	100	97.63	98.80	0h21m
BC (Mixed)	9.51	10.87	53.68	18.04	4h31m	3.10	0	0	0	4h27m	34.30	7.05	5.49	6.15	17h21m	59.18	100	40.82	57.98	5h11m	11.98	100	89.29	94.37	0h21m
BNC (Fixed)	48.01	46.76	48.91	47.82	4h31m	3.04	2.96	49.10	5.56	4h27m	70.62	42.91	19.11	26.27	17h21m	0	0	0	0	4h02m	2.86	100	97.23	98.59	0h21m
BNC (Grammar)	14.11	6.54	18.56	9.64	4h31m	4.62	0	0	0	4h27m	61.88	18.32	8.25	11.38	17h21m	0	0	0	0	4h15m	12.39	100	89.04	94.18	0h21m
CP (Variable)	49.24	49.83	50.76	50.29	4h31m	3.17	0	0	0	4h27m	82.43	24.17	12.35	16.42	17h21m	0	0	0	0	4h10m	15.58	100	86.78	92.91	0h21m
Average	34.05	22.80	34.38	25.16	4h31m	3.09	1.04	21.26	1.98	4h27m	62.76	25.99	12.70	16.96	17h21m	18.73	40.00	21.27	27.43	4h28m	9.06	100	91.99	93.77	0h21m
Code Search																									
BC (Fixed)	27.43	16.61	37.89	23.04	7h44m	7.67	5.25	40.47	9.26	7h42m	79.88	49.09	13.61	21.31	43h18m	0	0	0	0	50h17m	1.11	100	99.11	99.55	0h43m
BC (Mixed)	17.37	12.46	37.68	18.69	7h44m	9.71	6.97	41.78	12.06	7h42m	79.78	43.93	12.29	19.33	43h18m	0	0	0	0	51h15m	1.38	100	98.66	99.33	0h43m
BNC (Fixed)	8.63	6.10	37.79	10.52	7h44m	10.15	7.19	41.48	12.21	7h42m	79.97	42.82	12.29	19.19	43h18m	0	0	0	0	48h12m	3.10	100	97.06	98.51	0h43m
BNC (Grammar)	34.67	27.22	41.62	32.94	7h44m	7.76	7.66	49.67	13.36	7h42m	77.41	44.62	13.97	21.37	43h18m	0	0	0	0	49h10m	4.69	100	95.60	97.71	0h43m
CP (Variable)	45.93	21.56	27.39	24.10	7h44m	9.18	10.02	52.82	16.98	7h42m	80.66	35.12	11.34	17.20	43h18m	0	0	0	0	50h15m	20.31	100	83.36	90.97	0h43m
Average	26.75	16.79	36.47	21.86	7h44m	8.89	7.42	45.24	12.74	7h42m	79.54	43.12	12.70	19.68	43h18m	0	0	0	0	49h50m	6.12	100	94.76	97.21	0h43m
Code Code Repair																									
BC (Fixed)	30.07	98.61	76.58	86.33	24h48m	3.22	0	0	0	24h46m	75.09	46.54	14.70	22.49	31h26m	0	100	98.47	99.23	21h13m	0.53	100	100	100	0h5m
BC (Mixed)	30.84	13.85	29.92	18.77	24h48m	3.27	0	0	0	24h46m	79.31	45.12	13.53	21.05	31h26m	0	76.08	100	86.42	18h56m	1.44	100	98.57	99.28	0h5m
BNC (Fixed)	30.61	29.98	43.28	35.43	24h48m	3.17	2.22	42.51	4.21	24h46m	62.82	13.76	6.53	8.79	31h26m	0	0	0	0	19h46m	1.53	100	98.47	99.23	0h5m
BNC (Grammar)	30.59	99.84	76.66	86.83	24h48m	3.01	0	0	0	24h46m	65.56	28.67	11.20	16.15	31h26m	0	100	100	100	21h45m	2.67	100	97.42	98.69	0h5m
CP (Variable)	33.42	32.93	49.36	39.23	24h48m	3.15	3.33	51.41	6.21	24h46m	85.76	25.77	9.36	13.68	31h26m	0	0	0	0	18h04m	3.77	100	96.59	98.26	0h5m
Average	31.12	55.04	55.16	53.32	24h48m	3.16	1.11	18.78	2.08	24h46m	73.71	31.97	11.06	16.43	31h26m	0	55.22	60.00	57.28	19h57m	1.59	100	97.90	98.94	0h5m

* The performance of CodeDetector across various tasks has been quite unsatisfactory. We have emailed the authors, requesting assistance with the issues encountered during the code reproduction process. However, we have not yet received a response.

Performance of CodeDetector

Table XIV presents the overall performance of KillBadCode and baselines in detecting code poisoning. The results for AC, SS, ONION, and KillBadCode have already been analyzed in detail in the paper. For CodeDetector, it is almost unable to detect poisoned samples because the well-trained backdoored model does not react to individual tokens within the triggers, especially in the defect detection task and the code search task. We tried using different thresholds (including 0.1, 0.2, 0.3, and 0.4) when probing for trigger tokens, but it remains ineffective.