Appendix

Appendix A
Reliability Diagrams

A. *In Distribution*

Fig. A-1, Fig. 1, Fig. A-2 present the confidence distribution and reliability diagrams of different code models on source code classification task. Fig. A-3, Fig. 2, Fig. A-4 present the confidence distribution and reliability diagrams of different code models on clone detection, defect detection and exception type task, respectively.

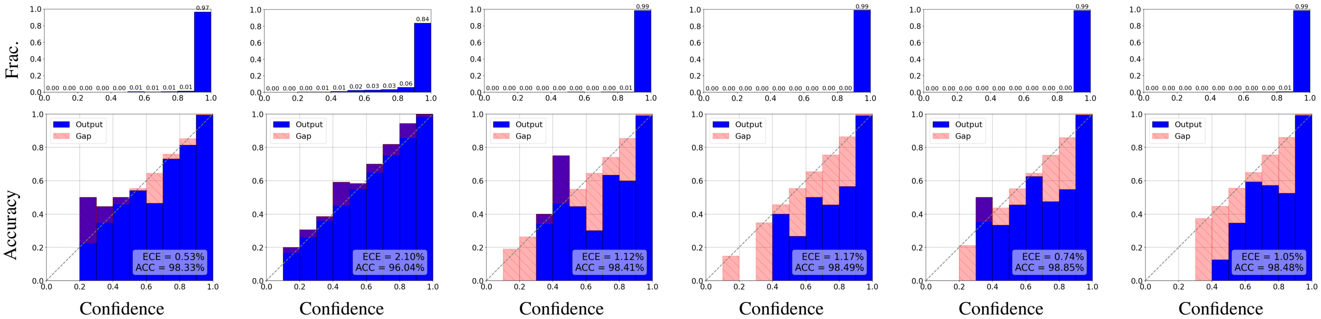
B. *Out of Distribution*

Fig. A-5, Fig. A-6 present the confidence distribution and reliability diagrams of different code models on CST-based and semantic-based OOD data of Java250, respectively.

Appendix B
Label Smoothing

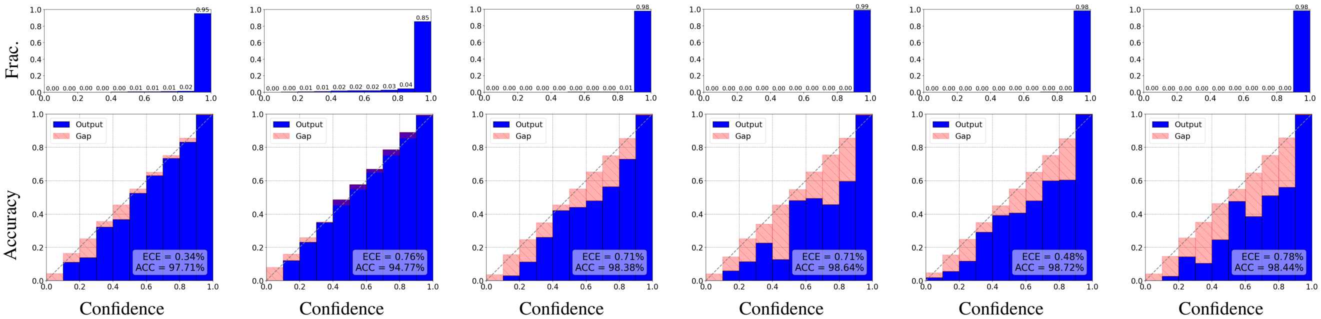
Table B-1 presents the results of label smoothing with more values($\alpha = 0, 0.1, 0.2, 0.3$) on different tasks in ID setting.

Table B-2 presents the results of label smoothing with more values($\alpha = 0.05$) on Java250 in different OOD settings.



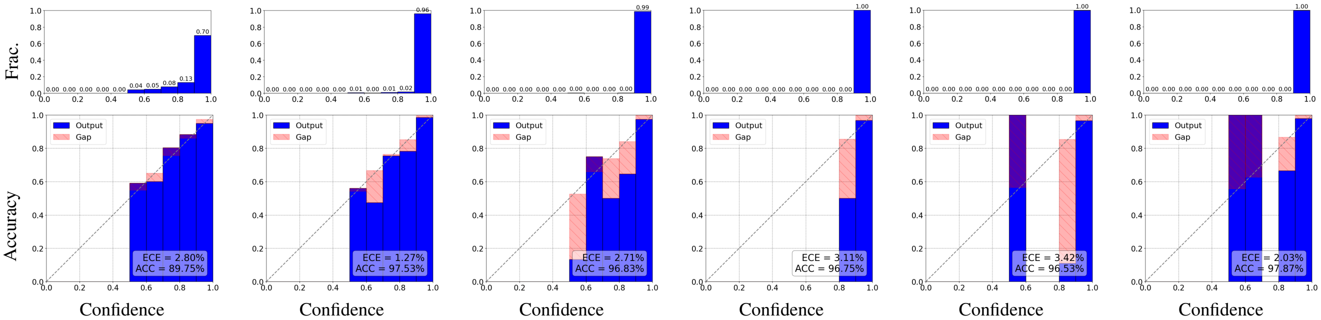
(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-1. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Code Classification (POJ104).



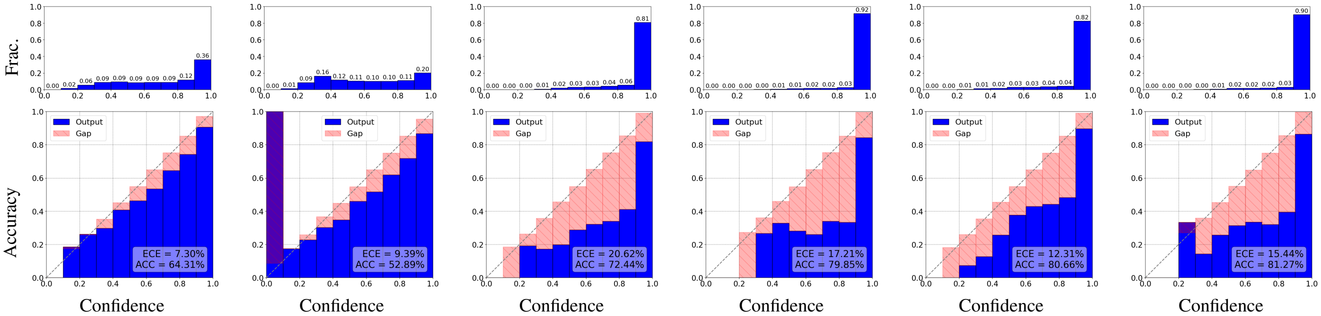
(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-2. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Code Classification (Python800).



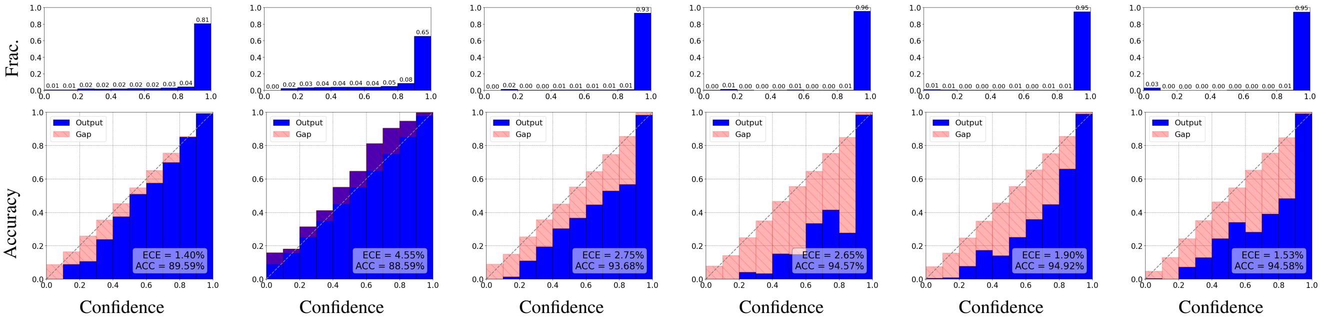
(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-3. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Code Clone Detection.



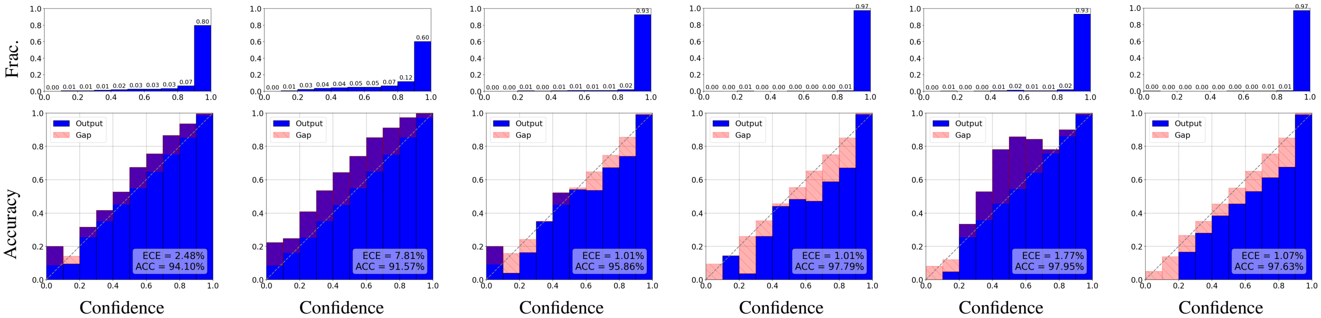
(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-4. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Exception Type.



(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-5. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Code Classification (Java250-OOD-CST).



(a) TextCNN (b) ASTNN (c) CodeBERTa (d) CodeBERT (e) GraphCodeBERT (f) CodeT5

Fig. A-6. Confidence distribution (top row) and reliability diagrams (bottom row) for different code models on Code Classification (Java250-OOD-semantic).

TABLE B-1
Experimental Results of Accuracy and ECE (%) with Different Label Smoothing Values

Task		Clone Detection				Defect Detection				Exception Type			
Models	Metric	0	0.1	0.2	0.3	0	0.1	0.2	0.3	0	0.1	0.2	0.3
TextCNN	Acc	89.75	90.15	90.45	90.77	62.63	63.40	63.43	63.69	64.31	65.00	64.93	65.13
	ECE	2.80	5.46 ↑	10.43↑	14.78 ↑	7.46	1.43	2.88	4.36	7.30	5.98	13.55 ↑	22.08 ↑
ASTNN	Acc	97.53	97.43	97.53	97.08	59.99	61.05	60.87	60.83	52.89	53.93	52.85	54.30
	ECE	1.27	4.94 ↑	9.34 ↑	13.55 ↑	9.76	6.56	2.69	1.77	9.39	0.94	7.53	11.74↑
CodeBERTa	Acc	96.82	96.90	96.78	96.55	62.52	61.57	61.97	62.37	72.44	73.78	73.68	73.44
	ECE	2.71	1.84	5.54 ↑	10.73↑	9.83	11.89 ↑	8.99	6.12	20.62	10.26	7.96	9.78
CodeBERT	Acc	96.75	97.03	96.82	96.50	63.25	62.23	63.80	63.21	79.85	79.41	79.40	79.86
	ECE	3.11	1.30	6.09 ↑	10.81 ↑	12.14	11.27	6.46	9.31	17.21	6.44	8.02	13.75
GraphCodeBERT	Acc	96.53	96.50	96.85	96.93	64.09	62.63	63.51	63.80	80.66	81.36	81.94	82.21
	ECE	3.42	1.83	6.25 ↑	10.67 ↑	8.82	15.67 ↑	4.85	3.13	12.31	5.63	9.06	14.76
CodeT5	Acc	97.88	97.12	96.75	97.25	63.91	62.15	62.45	63.29	81.27	80.69	81.34	80.14
	ECE	2.03	2.75 ↑	7.71 ↑	13.07 ↑	9.72	17.42 ↑	12.60 ↑	6.09	15.44	2.88	11.14	19.23↑

TABLE B-2

Experimental Results of OOD Datasets of Java250 with label smoothing=0.05

Dataset		ID	OOD-Token			OOD-CST			OOD-Semantic		
Models	Metrics	original	original	TS	LS	original	TS	LS	original	TS	LS
TextCNN	Acc	96.37	84.09	84.09	87.19	89.59	89.59	91.21	94.10	94.10	94.39
	ECE	0.38	2.84	2.99	15.49	1.40	3.40	12.39	2.48	5.64	14.53
ASTNN	Acc	96.52	80.21	80.21	82.44	88.59	88.59	91.27	91.57	91.57	92.03
	ECE	1.76	1.62	8.82	19.34	4.55	11.02	17.52	7.81	14.42	19.34
CodeBERTa	Acc	97.59	91.56	91.56	92.26	93.68	93.68	93.92	95.86	95.86	95.94
	ECE	0.90	3.68	0.60	2.32	2.75	0.54	2.34	1.01	3.50	3.60
CodeBERT	Acc	98.15	93.50	93.50	93.83	94.57	94.57	94.71	97.79	97.79	97.93
	ECE	0.87	3.43	1.24	2.67	2.65	0.97	3.30	1.01	1.40	2.89
GraphCodeBERT	Acc	98.29	93.99	93.99	94.15	94.92	94.92	94.97	97.95	97.95	97.89
	ECE	0.41	2.27	2.48	2.77	1.90	2.16	2.82	1.77	11.32	4.88
CodeT5	Acc	98.09	93.14	93.14	94.22	94.58	94.58	94.75	97.63	97.63	97.84
	ECE	0.61	2.86	1.34	4.25	1.53	0.61	3.51	1.07	0.49	4.84