

Table 1: Performance Comparison on Q-Error and P-Error.

Method	Q-Error			P-Error		
	50%	90%	99%	50%	90%	99%
Model A	1.12	3.45	10.1	1.01	2.89	9.95
Model B*	1.01	2.98	8.77	0.92	2.12	7.66
Model C	1.33	4.12	12.6	1.11	3.45	10.5

Note: * indicates supervised models. Bold indicates best result.

Table 2: Grouped Models with Multirow and Subtypes.

Graph-Based	GCN	0.98	4.55	0.76
	GAT	0.92	3.98	0.69
	GIN	1.10	4.88	0.81
Tree-Based	XGBoost	1.25	5.21	0.88
	LightGBM	1.14	5.00	0.83

Note: Values are in RMSE, lower is better.

Table 3: Ablation Results[†] with Footnotes.

Model Variant	Accuracy (%)
Full Model	91.2
w/o Attention	88.5
w/o Contrastive [‡]	86.9
w/o Fusion Layer	87.3

[†] Results based on average of 3 seeds. [‡] Contrastive refers to CL loss described in Sec 3.3.

Table 4: Extended Comparison across Multiple Metrics.

Method	Q-Error				P-Error			
	50%	75%	90%	99%	50%	75%	90%	99%
Model A	1.1	1.5	2.1	6.0	1.0	1.2	2.0	5.7
Model B	1.0	1.3	1.9	5.4	0.9	1.1	1.8	5.0
Model C	1.2	1.6	2.3	6.5	1.1	1.3	2.1	6.1
Model D	1.3	1.8	2.5	7.0	1.2	1.4	2.3	6.8
Model E*	1.1	1.4	2.0	5.9	1.0	1.1	1.9	5.6
Model F	1.4	1.9	2.7	7.3	1.3	1.5	2.5	7.0
Model G	1.2	1.5	2.2	6.2	1.1	1.2	2.0	5.8
Model H	1.3	1.7	2.4	6.7	1.2	1.4	2.2	6.4

Note: Bold indicates best result. * Indicates semi-supervised model.

Table 5: Grouped Evaluation with Multirow Headers.

Category	Model	F1 Score (%)			Accuracy (%)			Params (M)
		Easy	Med.	Hard	Easy	Med.	Hard	
Graph	GCN	91.2	85.6	79.4	93.1	87.5	81.2	12.3
	GAT	92.0	86.3	80.1	93.8	88.0	82.0	13.2
	GIN	90.5	84.7	78.2	92.3	86.1	80.0	11.7
	GraphSAGE	91.7	85.9	79.8	93.4	87.6	81.5	12.9
Tree	XGBoost	88.9	83.0	76.3	90.5	84.2	78.1	9.8
	LightGBM	89.1	83.5	76.9	90.7	84.8	78.6	9.5
	CatBoost	88.4	82.7	75.5	90.1	83.9	77.5	10.1

All models are evaluated on three difficulty subsets. F1 and Accuracy are macro-averaged.