Supplementary material to Robustness in Network Intrusion Detection with Adversarial Training and Ouf-of-Distribution

Anonymous Authors

This document presents supplementary material that supports our main paper. It includes tables displaying the results obtained for the metrics discussed in the paper across the relevant datasets. The data presented here pertains to the ensemble trained on unperturbed samples (TabNet) and the adversarial training (TabIDS). The graphs provided primarily correspond to the precision metric, with the exception of the ”Out-of-Distribution” section. In this section, the ROC-AUC metric is used for the TabIDS model, which is abbreviated as IDS, and for out-of-distribution (OOD) detection.

The sections in this supplement are organized according to adversarial attacks and out-of-distribution (OOD) scenarios, similar to the experiments section of the main paper. In the tables, blue-shaded cells indicate that the TabIDS model performed equally as well as the TabNet model. In contrast, yellow-shaded cells signify that the TabIDS model outperformed the TabNet model.

I. Projected Gradient Descent (PGD-100)

A. *CIC IDS2017*

TABLE I: PGD-100 attack against TabNet and TabIDS for binary classification on the CIC IDS2017 dataset.

Type	Norm	Metric	***Epsilon*** (ϵ)									
			0.008		**0.01**		**0.03**		**0.05**		**0.08**	
			Benign	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***
TabNet	ℓ_1	Precision	99.99	98.85	99.99	98.84	99.99	98.78	99.99	98.73	99.99	98.67
		Recall	99.76	99.95	99.76	99.94	99.75	99.94	99.74	99.94	99.73	99.94
		ROC-AUC	100.00	100.00	100.00	100.00	100.00	100.00	99.99	100.00	99.99	100.00
	ℓ_2	Precision	99.99	98.77	99.99	98.74	99.99	98.51	99.97	98.05	99.83	96.23
		Recall	99.75	99.94	99.74	99.94	99.69	99.93	99.59	99.83	99.21	99.16
		ROC-AUC	100.00	100.00	99.99	100.00	99.99	99.99	99.99	99.99	99.95	99.95
	ℓ_∞	Precision	99.98	98.34	99.98	98.12	99.37	86.50	92.39	48.56	77.24	1.22
		Recall	99.66	99.92	99.61	99.89	96.92	96.97	85.90	65.28	67.83	1.95
		ROC-AUC	99.99	99.99	99.99	99.99	99.44	99.47	87.17	87.17	22.53	22.45
TabIDS	ℓ_1	Precision	99.98	99.11	99.98	99.10	99.96	98.83	99.83	98.79	99.83	98.72
		Recall	99.82	99.89	99.82	99.89	99.76	99.82	99.75	99.17	99.74	99.16
		ROC-AUC	99.98	99.99	99.98	99.99	99.98	99.99	99.98	99.99	99.98	99.98
	ℓ_2	Precision	99.98	99.04	99.97	99.01	99.83	98.77	99.82	98.62	99.80	98.11
		Recall	99.80	99.88	99.80	99.87	99.75	99.16	99.72	99.13	99.61	99.04
		ROC-AUC	99.98	99.99	99.98	99.99	99.98	99.98	99.98	99.98	99.97	99.98
	ℓ_∞	Precision	99.93	98.79	99.88	98.38	99.80	98.04	99.80	98.09	99.79	98.01
		Recall	99.75	99.65	99.67	99.41	99.60	99.02	99.61	99.00	99.59	98.97
		ROC-AUC	99.98	99.99	99.98	99.98	99.97	99.98	99.97	99.98	99.97	99.97

B. *UNSW-NB15*

TABLE II: PGD-100 attack against TabNet and TabIDS for multiclass classification on the CIC IDS2017 dataset.

Type	**Norm**	**Metrics**	**Epsilon (ϵ)**														
			0.008			**0.01**			**0.03**			**0.05**			**0.08**		
			Benign	***DOS***	***DDOS***	***Benign***	***DOS***	***DDOS***	***Benign***	***DOS***	***DDOS***	***Benign***	***DOS***	***DDOS***	***Benign***	***DOS***	***DDOS***
TabNet	ℓ_1	Precision	99.98	99.30	99.99	99.97	99.30	99.99	99.97	99.26	99.99	99.97	99.24	99.99	99.97	99.21	99.99
		Recall	99.44	99.75	99.99	99.44	99.75	99.99	99.41	99.75	99.99	99.38	99.75	99.99	99.30	99.75	99.99
		ROC-AUC	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00
	ℓ_2	Precision	99.97	99.24	99.99	99.97	99.24	99.99	99.96	99.07	99.99	99.95	98.83	99.99	99.81	98.17	99.99
		Recall	99.41	99.75	99.99	99.39	99.75	99.99	99.17	99.72	99.97	98.83	99.64	99.94	98.23	99.40	99.83
		ROC-AUC	99.98	100.00	100.00	99.98	100.00	100.00	99.97	100.00	100.00	99.95	99.99	100.00	99.87	99.94	100.00
	ℓ_∞	Precision	99.96	98.94	99.99	99.95	98.74	99.99	99.55	90.49	99.96	97.21	63.13	99.80	85.39	15.56	94.45
		Recall	99.05	99.69	99.94	98.92	99.64	99.93	95.67	96.84	93.84	84.83	80.39	81.14	62.34	21.91	76.96
		ROC-AUC	99.96	99.99	100.00	99.95	99.99	100.00	99.38	99.00	99.99	94.16	96.12	99.37	59.82	63.23	90.39
TabIDS	ℓ_1	Precision	99.99	98.95	99.94	99.99	98.95	99.94	99.99	98.94	99.94	99.99	98.91	99.94	99.99	98.87	99.93
		Recall	97.81	99.93	100.00	97.79	99.93	100.00	97.56	99.91	100.00	97.30	99.91	99.99	97.06	99.90	99.95
		ROC-AUC	99.97	100.00	100.00	99.97	100.00	100.00	99.96	100.00	100.00	99.95	100.00	100.00	99.92	100.00	100.00
	ℓ_2	Precision	99.99	98.94	99.94	99.99	98.93	99.93	99.99	98.75	99.93	99.99	98.57	99.93	99.98	98.29	99.93
		Recall	97.62	99.92	100.00	97.55	99.91	100.00	96.72	99.89	99.96	96.04	99.86	99.98	96.06	99.85	99.99
		ROC-AUC	99.96	100.00	100.00	99.95	100.00	100.00	99.64	99.99	100.00	99.22	99.98	100.00	99.76	99.98	100.00
	ℓ_∞	Precision	99.99	98.66	99.92	99.99	98.60	99.93	99.97	97.49	99.92	99.94	94.87	99.91	99.73	83.52	99.91
		Recall	96.49	99.85	99.95	96.25	99.84	99.97	96.20	99.75	100.00	95.53	99.45	99.98	92.63	97.73	99.97
		ROC-AUC	99.75	99.99	100.00	99.47	99.99	100.00	99.68	99.98	100.00	99.59	99.93	100.00	99.03	99.63	100.00

TABLE IV: PGD-100 attack against TabNet and TabIDS for multiclass classification on the UNSW-NB15 dataset.

Type	**Norm**	**Metrics**	**Epsilon (ϵ)**														
			0.008			**0.01**			**0.03**			**0.05**			**0.08**		
			Normal	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***
TabNet	ℓ_1	Precision	100.00	80.10	100.00	100.00	80.41	100.00	100.00	78.12	100.00	100.00	75.71	100.00	100.00	58.23	99.85
		Recall	99.52	90.34	91.56	99.52	89.77	91.56	99.52	85.23	91.56	99.52	75.28	91.49	99.52	41.19	90.89
		ROC-AUC	99.96	99.98	99.97	99.96	99.98	99.97	99.95	99.97	99.97	99.95	99.95	99.97	99.95	99.88	99.96
	ℓ_2	Precision	100.00	79.74	100.00	100.00	78.41	100.00	100.00	53.02	99.93	100.00	24.85	99.71	100.00	9.15	94.98
		Recall	99.52	88.35	91.56	99.52	86.65	91.56	99.52	32.39	91.36	99.52	11.93	90.56	99.52	3.98	90.62
		ROC-AUC	99.95	99.98	99.97	99.95	99.98	99.97	99.94	99.84	99.96	99.93	99.67	99.95	99.90	99.44	99.93
	ℓ_∞	Precision	100.00	72.73	100.00	100.00	53.55	100.00	100.00	23.49	99.63	100.00	19.20	93.22	99.99	14.94	70.27
		Recall	99.52	70.45	91.49	99.52	32.10	91.42	99.52	11.08	89.89	99.52	6.82	87.77	99.53	6.53	24.20
		ROC-AUC	99.94	99.94	99.96	99.94	99.85	99.96	99.91	99.36	99.94	99.87	99.18	99.90	99.72	98.97	99.15
TabIDS	ℓ_1	Precision	100.00	50.97	97.86	100.00	48.08	97.65	100.00	40.00	97.55	100.00	34.03	94.11	100.00	30.10	87.41
		Recall	99.61	59.94	91.16	99.61	56.82	91.16	99.61	52.84	87.43	99.61	51.14	88.23	99.61	51.14	86.30
		ROC-AUC	99.93	99.88	99.95	99.93	99.87	99.95	99.93	99.85	99.94	99.92	99.80	99.93	99.92	99.78	99.93
	ℓ_2	Precision	100.00	43.55	96.52	100.00	38.48	95.76	100.00	28.24	85.59	100.00	22.96	77.08	100.00	18.70	75.20
		Recall	99.61	50.85	86.64	99.61	46.02	85.57	99.61	42.05	81.78	99.61	40.06	72.67	99.61	33.52	55.25
		ROC-AUC	99.93	99.86	99.94	99.93	99.84	99.93	99.91	99.74	99.90	99.89	99.71	99.87	99.88	99.65	99.83
	ℓ_∞	Precision	100.00	30.00	78.73	100.00	29.53	84.57	100.00	18.37	70.42	100.00	16.54	68.05	100.00	16.87	62.34
		Recall	99.61	39.20	69.15	99.61	38.92	63.03	99.61	30.68	45.28	99.61	30.11	41.36	99.61	31.82	42.15
		ROC-AUC	99.88	99.77	99.85	99.86	99.75	99.84	99.87	99.63	99.81	99.88	99.61	99.77	99.88	99.59	99.78

TABLE III: PGD-100 attack against TabNet and TabIDS for binary classification on the UNSW-NB15 dataset.

Type	**Norm**	**Metric**	***Epsilon*** (ϵ)									
			0.008		**0.01**		**0.03**		**0.05**		**0.08**	
			Benign	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***
TabNet	ℓ_1	Precision	100.00	85.99	100.00	85.99	100.00	85.94	100.00	85.92	100.00	85.90
		Recall	99.66	100.00	99.66	100.00	99.66	100.00	99.66	100.00	99.66	99.98
		ROC-AUC	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.97	99.97
	ℓ_2	Precision	100.00	85.94	100.00	85.92	100.00	85.90	100.00	85.85	99.99	85.77
		Recall	99.66	100.00	99.66	100.00	99.66	99.98	99.66	99.98	99.65	99.71
		ROC-AUC	99.98	99.98	99.98	99.98	99.97	99.97	99.95	99.95	99.93	99.93
	ℓ_∞	Precision	100.00	85.90	100.00	85.89	99.98	85.66	99.93	84.82	99.73	78.78
		Recall	99.66	99.98	99.66	99.98	99.65	99.05	99.64	96.49	99.51	87.24
		ROC-AUC	99.97	99.97	99.96	99.96	99.90	99.90	99.76	99.76	99.49	99.59
TabIDS	ℓ_1	Precision	99.94	60.59	99.94	60.47	99.93	59.71	99.92	59.45	99.91	59.24
		Recall	98.68	97.12	98.68	97.10	98.64	96.53	98.63	96.35	98.62	95.97
		ROC-AUC	99.87	99.87	99.87	99.87	99.85	99.84	99.83	99.83	99.79	99.79
	ℓ_2	Precision	99.94	60.34	99.93	60.27	99.92	59.23	99.91	59.12	99.92	58.91
		Recall	98.67	96.96	98.67	96.89	98.62	96.02	98.62	95.79	98.60	96.02
		ROC-AUC	99.86	99.86	99.85	99.85	99.82	99.82	99.78	99.78	99.73	99.73
	ℓ_∞	Precision	99.92	56.82	99.92	56.97	99.91	58.58	99.90	47.59	99.88	51.82
		Recall	98.47	96.24	98.49	96.04	98.59	95.66	97.81	95.41	98.17	**94.42**
		ROC-AUC	99.79	99.79	99.72	99.72	99.67	99.66	99.59	99.59	99.58	99.58

II. Carlini-Wagner (CW-2)

A. *CIC IDS2017*

TABLE V: CW-2 attack against TabNet and TabIDS for binary classification on the CIC IDS2017 dataset.

Type	**Metric**	**Confidence**									
		0.0		**0.2**		**0.5**		**0.8**		**1.0**	
		Benign	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***
TabNet	Precision	99.99	98.69	99.99	98.70	99.99	98.71	99.99	98.74	99.99	98.74
	Recall	99.73	99.95	99.73	99.95	99.73	99.95	99.74	99.95	99.74	99.95
	ROC-AUC	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
TabIDS	Precision	99.98	99.07	99.98	99.09	99.98	99.11	99.98	99.11	99.98	99.11
	Recall	99.81	99.89	99.81	99.89	99.82	99.89	99.82	99.89	99.82	99.89
	ROC-AUC	99.98	99.99	99.98	99.99	99.98	99.99	99.98	99.99	99.98	99.99

TABLE VI: CW-2 attack against TabNet and TabIDS for multiclass classification on the CIC IDS2017 dataset.

Type	**Metric**	**Confidence**														
		0.0			**0.2**			**0.5**			**0.8**			**1.0**		
		Benign	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***
TabNet	Precision	99.99	99.23	99.99	99.99	99.25	99.99	99.99	99.26	99.99	99.99	99.28	99.99	99.99	99.29	99.99
	Recall	99.36	99.75	99.99	99.38	99.75	99.99	99.39	99.75	99.99	99.41	99.75	99.99	99.42	99.76	99.99
	ROC-AUC	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00	99.98	100.00	100.00
TabIDS	Precision	99.99	98.92	99.94	99.99	98.93	99.94	99.99	98.95	99.94	99.99	98.96	99.94	99.99	98.97	99.94
	Recall	97.79	99.93	100.00	97.85	99.93	100.00	97.86	99.93	100.00	97.87	99.93	100.00	97.88	99.93	100.00
	ROC-AUC	99.97	100.00	100.00	99.97	100.00	100.00	99.97	100.00	100.00	99.97	100.00	100.00	99.97	100.00	100.00

B. *UNSW-NB15*

TABLE VII: CW-2 attack against TabNet and TabIDS for binary classification on the UNSW-NB15 dataset.

Type	**Metric**	**Confidence**									
		0.0		**0.2**		**0.5**		**0.8**		**1.0**	
		Normal	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***
TabNet	Precision	100.00	84.07	100.00	84.15	100.00	84.31	100.00	84.40	100.00	84.51
	Recall	99.61	99.86	99.61	99.86	99.61	99.86	99.62	99.89	99.62	99.93
	ROC-AUC	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98
TabIDS	Precision	99.87	4.99	99.88	5.12	99.89	5.33	99.90	5.57	99.90	5.73
	Recall	61.85	96.11	62.81	96.26	64.20	96.71	65.79	96.69	66.82	96.76
	ROC-AUC	98.41	98.41	98.49	98.49	98.68	98.68	98.70	98.70	98.75	98.75

TABLE VIII: CW-2 attack against TabNet and TabIDS for multiclass classification on the UNSW-NB15 dataset.

Type	**Metric**	**Confidence**														
		0.0			**0.2**			**0.5**			**0.8**			**1.0**		
		Normal	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***
TabNet	Precision	100.00	32.34	100.00	100.00	33.53	100.00	100.00	34.13	100.00	100.00	34.13	100.00	100.00	34.32	100.00
	Recall	70.09	15.34	84.77	70.12	16.19	85.44	70.20	16.19	86.44	70.30	16.19	87.50	70.36	16.48	87.90
	ROC-AUC	98.92	98.34	99.03	98.91	98.26	99.02	98.91	98.22	99.01	98.90	98.20	99.00	98.89	98.16	99.00
TabIDS	Precision	100.00	13.78	97.46	100.00	14.74	97.82	100.00	18.81	97.75	100.00	21.23	97.75	100.00	24.63	98.03
	Recall	99.59	12.22	89.43	99.59	13.07	89.43	99.59	17.05	89.56	99.59	19.60	89.56	99.59	23.58	89.56
	ROC-AUC	99.92	99.82	99.94	99.92	99.81	99.94	99.92	99.80	99.94	99.92	99.80	99.94	99.92	99.80	99.94

III. DeepFool (DF-2)

A. *CIC IDS2017*

TABLE IX: DF-2 attack against TabNet and TabIDS for binary classification on the CIC IDS2017 dataset.

Type	**Metric**	**Confidence**									
		0.001		**0.002**		**0.005**		**0.008**		**0.009**	
		Benign	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***	***Benign***	***Attacks***
TabNet	Precision	77.28	0.75	77.26	0.74	77.19	0.74	77.12	0.73	77.09	0.72
	Recall	68.51	1.16	68.45	1.16	68.17	1.16	67.89	1.15	67.78	1.15
	ROC-AUC	1.74	1.72	1.73	1.72	1.72	1.70	1.70	1.68	1.69	1.68
TabIDS	Precision	94.48	60.21	94.47	60.16	94.44	60.02	94.41	59.89	94.40	59.84
	Recall	90.00	74.21	89.99	74.17	89.95	73.99	89.92	73.86	89.91	73.81
	ROC-AUC	92.98	92.98	92.96	92.96	92.89	92.89	92.82	92.82	92.80	92.80

TABLE X: DF-2 attack against TabNet and TabIDS for multiclass classification on the CIC IDS2017 dataset.

Type	**Metric**	**Epsilon (ϵ)**														
		0.001			**0.002**			**0.005**			**0.008**			**0.009**		
		Benign	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***	***Benign***	***DoS***	***DDoS***
TabNet	Precision	66.67	0.18	1.95	66.52	0.17	1.90	66.11	0.17	1.70	65.66	0.17	1.54	65.49	0.16	1.48
	Recall	25.43	0.41	1.12	25.28	0.41	1.10	24.83	0.40	1.00	24.35	0.39	0.92	24.17	0.38	0.89
	ROC-AUC	27.69	7.11	93.96	27.63	7.08	93.92	27.44	7.00	93.80	27.27	6.92	93.68	27.21	6.90	93.64
TabIDS	Precision	96.87	69.88	95.19	96.87	69.86	95.18	96.86	69.79	95.14	96.86	69.72	95.12	96.85	69.68	95.11
	Recall	92.12	69.46	92.51	92.11	69.44	92.48	92.09	69.38	92.46	92.05	69.32	92.41	92.04	69.28	92.38
	ROC-AUC	95.73	95.62	99.36	95.72	95.61	99.35	95.71	95.60	99.35	95.71	95.58	99.34	95.70	95.57	99.34

B. UNSW-NB15

TABLE XI: DF-2 attack against TabNet and TabIDS for binary classification on the UNSW-NB15 dataset.

Type	Metric	Confidence									
		0.001		0.002		0.005		0.008		0.009	
		Normal	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***	***Normal***	***Attacks***
TabNet	Precision	99.42	2.76	99.41	2.75	99.41	2.74	99.40	2.73	99.40	2.73
	Recall	33.34	90.61	33.24	90.61	32.95	90.55	32.69	90.50	32.61	90.57
	ROC-AUC	78.03	78.03	77.99	77.99	77.90	77.90	77.79	77.79	77.76	77.76
TabIDS	Precision	99.49	3.44	99.48	3.42	99.47	3.35	99.45	3.29	99.45	3.27
	Recall	48.52	88.03	48.17	87.96	47.14	87.94	46.09	87.87	45.74	87.85
	ROC-AUC	60.38	60.37	60.06	60.06	59.11	59.11	58.18	58.17	57.87	57.86

TABLE XII: DF-2 attack against TabNet and TabIDS for multiclass classification on the UNSW-NB15 dataset.

Type	Metric	Epsilon (ϵ)														
		0.001			0.002			0.005			0.008			0.009		
		Normal	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***	***Normal***	***Recon.***	***Generic***
TabNet	Precision	100.00	60.98	95.42	100.00	60.73	95.40	100.00	60.98	95.36	100.00	60.98	95.32	100.00	60.98	95.30
	Recall	99.42	42.61	63.70	99.40	42.61	63.43	99.33	42.61	62.90	99.24	42.61	62.23	99.20	42.61	62.03
	ROC-AUC	99.93	99.78	99.90	99.93	99.78	99.90	99.93	99.77	99.89	99.91	99.77	99.89	99.91	99.77	99.89
TabIDS	Precision	100.00	17.92	66.02	100.00	17.95	66.02	100.00	17.95	65.73	100.00	17.90	66.89	100.00	17.87	66.89
	Recall	86.21	32.39	20.28	86.20	32.39	20.28	86.13	32.39	20.15	86.07	32.39	20.28	86.05	32.39	20.28
	ROC-AUC	98.14	99.47	99.34	98.13	99.47	99.34	98.11	99.46	99.33	98.09	99.46	99.33	98.08	99.46	99.33

IV. Boundary, HopSkipJump, SignOPT

A. CIC IDS2017

TABLE XIII: Black-Box attacks against TabNet and TabIDS for binary classification on the CIC IDS2017 dataset.

Type	Attacks	Metrics	Label	
			Benign	**Attacks**
TabNet	Boundary	Precision	83.02	16.83
		Recall	67.71	32.07
		ROC-AUC	52.23	58.04
	HopSkipJump	Precision	83.02	16.82
		Recall	67.81	31.93
		ROC-AUC	73.71	73.92
	SignOPT	Precision	85.73	22.51
		Recall	69.88	42.93
		ROC-AUC	80.60	81.02
TabIDS	Boundary	Precision	83.10	17.09
		Recall	83.88	16.30
		ROC-AUC	82.08	83.52
	HopSkipJump	Precision	83.10	17.07
		Recall	83.80	16.36
		ROC-AUC	86.37	86.38
	SignOPT	Precision	86.98	40.52
		Recall	89.82	34.01
		ROC-AUC	93.21	93.22

TABLE XIV: Black-Box attacks against TabNet and TabIDS for multiclass classification on the CIC IDS2017 dataset.

Type	**Attacks**	**Metrics**	**Label**		
			Benign	***DoS***	***DDoS***
TabNet	Boundary	Precision	83.10	10.44	10.28
		Recall	57.43	4.47	0.06
		ROC-AUC	49.91	45.61	46.71
	HopSkipJump	Precision	79.79	13.37	4.71
		Recall	57.59	4.67	0.02
		ROC-AUC	63.25	95.63	99.87
	SignOPT	Precision	81.84	18.73	21.05
		Recall	60.70	11.31	1.16
		ROC-AUC	70.63	93.96	99.22
TabIDS	Boundary	Precision	83.07	10.01	5.73
		Recall	72.21	11.07	3.86
		ROC-AUC	75.44	77.64	76.64
	HopSkipJump	Precision	83.17	10.56	8.27
		Recall	72.13	11.66	3.82
		ROC-AUC	82.96	94.62	98.68
	SignOPT	Precision	86.65	23.13	34.06
		Recall	74.11	19.78	17.69
		ROC-AUC	88.30	96.33	98.83

B. *UNSW-NB15*

TABLE XV: Black-Box Attacks against TabNet and TabIDS for binary classification on the UNSW-NB15 Dataset.

Type	**Attacks**	**Metrics**	**Label**	
			Normal	**Attacks**
TabNet	Boundary	Precision	97.95	2.01
		Recall	89.46	10.38
		ROC-AUC	51.94	64.94
	HopSkipJump	Precision	97.95	1.95
		Recall	89.45	10.06
		ROC-AUC	90.36	90.37
	SignOPT	Precision	98.50	6.71
		Recall	90.01	34.46
		ROC-AUC	93.33	93.36
TabIDS	Boundary	Precision	97.96	2.07
		Recall	82.97	17.29
		ROC-AUC	56.65	73.07
	HopSkipJump	Precision	97.97	2.10
		Recall	83.08	17.38
		ROC-AUC	85.89	85.89
	SignOPT	Precision	98.72	6.85
		Recall	87.00	45.78
		ROC-AUC	92.90	92.90

TABLE XVI: Black-Box attacks against TabNet and TabIDS for multiclass classification on the UNSW-NB15 dataset.

Type	**Attacks**	**Metrics**	**Label**		
			Normal	***Recon.***	***Generic***
TabNet	Boundary	Precision	97.96	0.00	0.00
		Recall	93.37	0.00	0.00
		ROC-AUC	50.94	50.16	50.85
	HopSkipJump	Precision	99.33	6.12	1.47
		Recall	93.26	1.70	0.07
		ROC-AUC	96.19	99.46	99.27
	SignOPT	Precision	99.60	1.80	1.14
		Recall	93.64	1.42	0.07
		ROC-AUC	98.17	99.46	99.09
TabIDS	Boundary	Precision	97.94	0.00	0.39
		Recall	51.97	0.00	0.20
		ROC-AUC	64.09	63.20	62.57
	HopSkipJump	Precision	99.50	6.27	5.07
		Recall	51.87	23.01	0.47
		ROC-AUC	96.08	97.73	97.13
	SignOPT	Precision	99.73	2.01	15.26
		Recall	68.50	16.19	10.24
		ROC-AUC	98.12	97.95	97.69

V. Out-of-Distribution Detection (OOD)

A. *CIC IDS2017*

TABLE XVIII: ROC-AUC metrics for TabIDS and OOD in multiclass classification and CIC IDS2017 dataset.

Attack	**Parameter**	**TabIDS**	**OOD**
PGD-100 ℓ_2	0.008	99.99	48.8
	0.01	99.95	48.9
	0.03	99.77	52.1
	0.05	99.79	58.6
	0.08	99.80	67.2
PGD-100 ℓ_∞	0.008	99.99	48.9
	0.01	99.79	49.9
	0.03	99.62	66.5
	0.05	99.55	77.5
	0.08	99.58	88.2
DF-2	0.001	95.73	90.1
	0.002	97.77	90.1
	0.005	97.79	90.1
	0.008	98.10	90.1
	0.009	98.37	90.1
CW-2	0	99.97	48.7
	0.2	99.99	48.7
	0.5	99.99	48.7
	0.8	99.99	48.7
	1.0	99.94	48.7
Boundary		75.44	98.5
HopSkipJump		82.95	98.5
SignOPT		88.30	98.4

B. UNSW-NB15

TABLE XIX: ROC-AUC metrics for TabIDS and OOD in binary classification and UNSW-NB15 dataset.

Attack	**Parameter**	**TabIDS**	**OOD**
PGD-100 ℓ_2	0.008	99.86	49.5
	0.01	99.85	49.8
	0.03	99.82	53.9
	0.05	99.78	57.8
	0.08	99.73	64.5
PGD-100 ℓ_∞	0.008	99.79	60.6
	0.01	99.72	62.2
	0.03	99.66	86.9
	0.05	99.59	87.8
	0.08	99.58	87.7
DF-2	0.001	78.03	73.7
	0.002	77.99	73.8
	0.005	77.90	73.9
	0.008	77.79	74.1
	0.009	77.76	74.2
CW-2	0.0	98.41	45.5
	0.2	98.49	45.4
	0.5	98.68	45.2
	0.8	98.70	45.2
	1.0	98.75	45.2
Boundary		73.07	100
Hopskipjump		85.89	100
Signopt		92.90	100

TABLE XX: ROC-AUC metrics for TabIDS and OOD in binary classification and UNSW-NB15 dataset.

Attack	**Parameter**	**TabIDS**	**OOD**
PGD-100 ℓ_2	0.008	99.77	49.7
	0.01	99.80	50
	0.03	99.79	53
	0.05	99.76	55.9
	0.08	99.71	59.2
PGD-100 ℓ_∞	0.008	99.68	58.4
	0.01	99.71	63
	0.03	99.69	85.6
	0.05	99.66	88
	0.08	99.62	89.5
DF-2	0.001	98.14	91.1
	0.002	93.98	91.1
	0.005	94.00	91.1
	0.008	95.07	91.1
	0.009	95.79	91.1
CW	0.0	99.92	49
	0.2	99.84	49
	0.5	99.77	49
	0.8	99.81	49
	1.0	99.79	49
Boundary		64.09	98.9
Hopskipjump		96.08	98.9
Signopt		98.12	98.9