

Success Rates per Confidence Percentile ( $\uparrow$ )

Attack Name

No Attack	50%	4%	0%	0%	0%	0%	0%
BR5V	50%	37%	30%	26%	23%	21%	19%
BRS	50%	42%	37%	33%	30%	28%	26%
1VITAt	78%	57%	53%	51%	50%	48%	46%
3VITAt	82%	64%	59%	56%	54%	53%	51%
5VITAt	83%	66%	61%	58%	56%	54%	52%
fullVITAt	92%	82%	76%	71%	67%	65%	62%
FGSMt	99%	98%	95%	91%	86%	82%	76%
BIMt	99%	98%	96%	92%	86%	82%	77%
	50	60	70	80	90	95	99

Model Prediction Confidence Percentile