Table 1: Targeted attack effectiveness (%) comparing four settings: white-box backdoor (gradient-access, manual annotations), gray-box backdoor (gradient-access, annotation-free), our black-box backdoor (gradient-free, annotation-free), and a baseline black-box adversarial attack (gradient-free, annotation-free, no backdoor training).

Dataset	Model	ASR				CA			
		White-box Backdoor (TSBA) [9]	Gray-box Backdoor	Black-box Backdoor (Ours)	Black-box Adv. [13]	White-box Backdoor (TSBA) [9]	Gray-box Backdoor	Black-box Backdoor (Ours)	Black-box Adv. [13]
iAWE	FCN	99.5	92.5	83.3	55.4	99.7	99.7	90.3	99.9
	Resnet	99.1	96.0	92.8	50.8	98.2	98.2	94.9	98.3
	Transformer	96.2	90.1	87.9	48.8	96.4	96.4	95.1	96.2
MotionSense	FCN	99.3	75.9	73.2	51.5	99.3	99.3	99.0	99.2
	Resnet	99.1	77.6	71.1	52.5	99.4	99.4	98.6	98.6
	Transformer	97.6	65.2	62.6	35.2	97.5	97.3	97.2	97.6

Note: Both the gray-box and black-box settings use our proposed pipeline; the gray-box variant additionally assumes direct access to the target model's gradients.