ributions of BSD at 1.0 Detection Threshold L2CarliniWagnerAttack on C 100% 80% Classification accuracy 60% ISD 10 Accuracy JSD 40 Accuracy 40% JSD 100 Accuracy MagNet Accuracy Total Accuracy Class Divergence Detector 20% Total JSD detector Distance Based Detections 0% 20 60 40 80 100

Confidence in L2CarliniWagnerAttack