NEW EAD Attack against BSD at 2.0% Detection Threshold MNIST 100% 80% 60% 40% MagNet 20% BSD with JSD BSD w/o JSD 0% 10 30 20 35

Confidence in EAD attack

Classification accuracy