Classes Confidence CWL2 Attack against BSD at 0.0% Detection Threshold

- MagNet
- BSD with JSD
- BSD with Confidence
- BSD w/o Confidence and JSD

X-axis: Confidence in CWL2 attack
Y-axis: Classification accuracy