ML in Cyber Security

Lab 4 - Backdoor Attacks

Shreya Agarwal

sa6981@nyu.edu

Table with the accuracy on clean test data and the attack success rate (on backdoored test data) as a function of the fraction of channels pruned (X)

Channel Index	Accuracy
0	98.648
26	98.648
27	98.648
30	98.648
31	98.648
33	98.648
34	98.648
36	98.648
37	98.648
38	98.648
25	98.648
39	98.648
41	98.648
44	98.648

```
| 45            | 98.648   |

| 47            | 98.648   |

| 48            | 98.648   |

| 49            | 98.648   |

| 50            | 98.648   |

| 53            | 98.648   |

| 55            | 98.648   |

| 40            | 98.648   |

| 24            | 98.648   |

| 59            | 98.648   |

| 9             | 98.648   |

| 2             | 98.648   |

| 12            | 98.648   |

| 13            | 98.648   |

| 17            | 98.648   |

| 14            | 98.648   |

| 15            | 98.648   |

| 23            | 98.648   |

| 6             | 98.648   |

| 51            | 98.640   |

| 32            | 98.640   |

| 22            | 98.632   |

| 21            | 98.658   |

| 20            | 98.649   |
```

19	98.606
43	98.571
58	98.536
3	98.190
42	97.653
1	97.506
29	95.756
16	95.202
56	94.717
46	92.093
5	91.496
8	91.019
11	89.175
54	84.438
10	76.487
28	54.863
35	27.089
18	13.874
4	7.101
7	1.550