| penality | without\_res | vpp | vpn | accuracy |
| --- | --- | --- | --- | --- |
| 0.7 | 6 | 92.8% [88.8-95.5%] | 96.9% [95-98.1%] | 95.6% [93.9-96.9%] |
| 0.8 | 6 | 93.2% [89.2-95.7%] | 96.9% [95-98.1%] | 95.7% [94-97%] |
| 0.9 | 1 | 92.8% [88.7-95.4%] | 96.7% [94.8-98%] | 95.5% [93.8-96.8%] |
| 1.0 | 0 | 92.7% [88.7-95.4%] | 96.6% [94.6-97.8%] | 95.4% [93.6-96.7%] |
| 1.1 | 0 | 91.7% [87.5-94.5%] | 97.1% [95.3-98.2%] | 95.4% [93.6-96.7%] |
| 1.2 | 0 | 93.6% [89.7-96.1%] | 96.9% [95.1-98.1%] | 95.9% [94.2-97.1%] |
| 1.3 | 0 | 92.3% [88.2-95.1%] | 96.4% [94.4-97.7%] | 95.1% [93.3-96.4%] |
| 1.4 | 13 | 93.6% [89.5-96.1%] | 94.9% [92.6-96.4%] | 94.5% [92.6-95.9%] |

| temp | without\_res | vpp | vpn | accuracy |
| --- | --- | --- | --- | --- |
| 0.1 | 0 | 92.7% [88.7-95.4%] | 96.6% [94.6-97.8%] | 95.4% [93.6-96.7%] |
| 0.2 | 0 | 93.1% [89.1-95.7%] | 96.6% [94.6-97.8%] | 95.5% [93.8-96.8%] |
| 0.3 | 0 | 92.3% [88.2-95.1%] | 96.6% [94.6-97.8%] | 95.2% [93.5-96.5%] |
| 0.4 | 1 | 92.8% [88.7-95.4%] | 96.7% [94.8-98%] | 95.5% [93.8-96.8%] |
| 0.5 | 0 | 93.5% [89.5-96%] | 96.2% [94.2-97.5%] | 95.4% [93.6-96.7%] |
| 0.6 | 0 | 92.4% [88.3-95.1%] | 96.9% [95.1-98.1%] | 95.5% [93.8-96.8%] |
| 0.7 | 0 | 94.3% [90.5-96.6%] | 96.2% [94.2-97.5%] | 95.6% [93.9-96.9%] |
| 0.8 | 0 | 91.6% [87.4-94.5%] | 96.7% [94.8-97.9%] | 95.1% [93.3-96.4%] |

| prompt | without\_res | vpp | vpn | accuracy |
| --- | --- | --- | --- | --- |
| 0 | 0 | 90.2% [85.8-93.3%] | 97.3% [95.5-98.4%] | 95% [93.2-96.3%] |
| 1 | 0 | 81.5% [76.5-85.7%] | 97.9% [96.2-98.9%] | 91.9% [89.8-93.7%] |
| 2 | 0 | 92.7% [88.7-95.4%] | 96.6% [94.6-97.8%] | 95.4% [93.6-96.7%] |
| 3 | 65 | 63.5% [58.3-68.4%] | 99.4% [97.9-99.8%] | 81.5% [78.4-84.2%] |
| 4 | 13 | 84.6% [79.7-88.6%] | 97.6% [95.8-98.6%] | 93.2% [91.1-94.8%] |
| 5 | 0 | 89% [83.9-92.6%] | 89.8% [87-92%] | 89.6% [87.2-91.5%] |
| 6 | 0 | 87% [82.2-90.6%] | 95.9% [93.8-97.3%] | 93% [91-94.6%] |