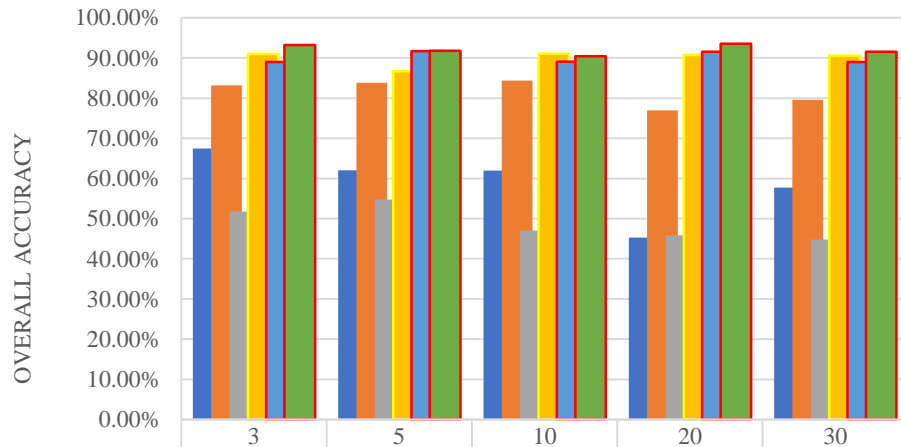


(a)

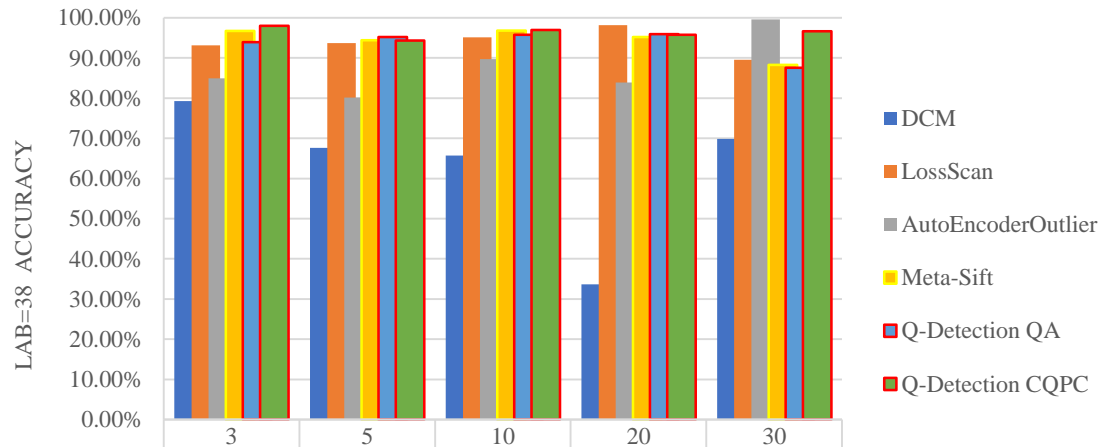
OVERALL ACCURACY UNDER TARGET LABEL-FLIPPING ATTACK



POISON RATE

(b)

ACCURACY OF LAB = 38 UNDER TARGET LABEL-FLIPPING ATTACK



POISON RATE

| | | | | | |
|--------------------|--------|--------|--------|--------|--------|
| DCM | 67.49% | 62.02% | 61.95% | 45.30% | 57.73% |
| LossScan | 83.17% | 83.80% | 84.34% | 76.97% | 79.58% |
| AutoEncoderOutlier | 51.77% | 54.82% | 47.09% | 45.84% | 44.82% |
| Meta-Sift | 90.95% | 86.67% | 91.09% | 90.75% | 90.58% |
| Q-Detection QA | 88.99% | 91.68% | 89.06% | 91.52% | 88.95% |
| Q-Detection CQPC | 93.24% | 91.74% | 90.43% | 93.53% | 91.51% |

| | | | | | |
|--------------------|--------|--------|--------|--------|--------|
| DCM | 79.28% | 67.59% | 65.68% | 33.62% | 69.86% |
| LossScan | 93.14% | 93.72% | 95.09% | 98.13% | 89.57% |
| AutoEncoderOutlier | 84.93% | 80.14% | 89.71% | 83.91% | 99.57% |
| Meta-Sift | 96.74% | 94.38% | 96.84% | 95.24% | 88.26% |
| Q-Detection QA | 93.96% | 95.21% | 95.79% | 95.92% | 87.53% |
| Q-Detection CQPC | 97.97% | 94.34% | 96.95% | 95.80% | 96.62% |

- DCM
- LossScan
- AutoEncoderOutlier
- Meta-Sift
- Q-Detection QA
- Q-Detection CQPC