# Evaluating Imperceptibility of White-Box Adversarial Attacks on Tabular Data

No Author Given

No Institute Given

## Supplementary Materials

### A  Evaluation Results on Performance

The evaluation results on performance can be found in Table 1 – 5.

### B  Evaluation Results on Imperceptibility

The evaluation results on imperceptibility can be found in Table 6 – 10.

| Dataset | ML models | Natural Accuracy | Attack methods | Robust Accuracy | Success Rate |
|---|---|---|---|---|---|
| Adult | LR | 0.8524 | DeepFool | 0.1731 | 0.8269 |
| | | | C&W L2 | 0.3014 | 0.6986 |
| | | | C&W Linf | 0.8445 | 0.1555 |
| | | | FGSM L1 | 0.7881 | 0.2119 |
| | | | FGSM L2 | 0.3232 | 0.6768 |
| | | | FGSM Linf | 0.1505 | 0.8495 |
| | | | BIM | 0.1505 | 0.8495 |
| | | | MIM | 0.1505 | 0.8495 |
| | | | PGD L1 | 0.7068 | 0.2932 |
| | | | PGD L2 | 0.2590 | 0.7410 |
| | | | PGD Linf | 0.1505 | 0.8495 |
| | Linear SVC | 0.8532 | DeepFool | 0.1468 | 0.8532 |
| | | | C&W L2 | 0.8524 | 0.1476 |
| | | | C&W Linf | 0.8456 | 0.1544 |
| | | | FGSM L1 | 0.6860 | 0.3140 |
| | | | FGSM L2 | 0.1912 | 0.8088 |
| | | | FGSM Linf | 0.1525 | 0.8475 |
| | | | BIM | 0.1525 | 0.8475 |
| | | | MIM | 0.1525 | 0.8475 |
| | | | PGD L1 | 0.4664 | 0.5336 |
| | | | PGD L2 | 0.1555 | 0.8445 |
| | | | PGD Linf | 0.1525 | 0.8475 |
| | MLP | 0.8521 | DeepFool | 0.1549 | 0.8451 |
| | | | C&W L2 | 0.5481 | 0.4519 |
| | | | C&W Linf | 0.8444 | 0.1556 |
| | | | FGSM L1 | 0.7503 | 0.2497 |
| | | | FGSM L2 | 0.2608 | 0.7392 |
| | | | FGSM Linf | 0.1535 | 0.8465 |
| | | | BIM | 0.1606 | 0.8394 |
| | | | MIM | 0.1485 | 0.8515 |
| | | | PGD L1 | 0.4876 | 0.5124 |
| | | | PGD L2 | 0.1719 | 0.8281 |
| | | | PGD Linf | 0.1485 | 0.8515 |

**Table 1.** Performance benchmark results for dataset: Adult.

| Dataset | ML models | Natural Accuracy | Attack methods | Robust Accuracy | Success Rate |
|---------|-----------|------------------|----------------|-----------------|--------------|
| German | LR | 0.8125 | DeepFool | 0.1875 | 0.8125 |
| | | | C&W L2 | 0.2760 | 0.7240 |
| | | | C&W Linf | 0.8021 | 0.1979 |
| | | | FGSM L1 | 0.8021 | 0.1979 |
| | | | FGSM L2 | 0.7083 | 0.2917 |
| | | | FGSM Linf | 0.1927 | 0.8073 |
| | | | BIM | 0.1927 | 0.8073 |
| | | | MIM | 0.1927 | 0.8073 |
| | | | PGD L1 | 0.8073 | 0.1927 |
| | | | PGD L2 | 0.6406 | 0.3594 |
| | | | PGD Linf | 0.1927 | 0.8073 |
| | Linear SVC | 0.8125 | DeepFool | 0.1875 | 0.8125 |
| | | | C&W L2 | 0.8177 | 0.1823 |
| | | | C&W Linf | 0.8073 | 0.1927 |
| | | | FGSM L1 | 0.8125 | 0.1875 |
| | | | FGSM L2 | 0.6667 | 0.3333 |
| | | | FGSM Linf | 0.1927 | 0.8073 |
| | | | BIM | 0.1927 | 0.8073 |
| | | | MIM | 0.1927 | 0.8073 |
| | | | PGD L1 | 0.8073 | 0.1927 |
| | | | PGD L2 | 0.6042 | 0.3958 |
| | | | PGD Linf | 0.1927 | 0.8073 |
| | MLP | 0.7969 | DeepFool | 0.1927 | 0.8073 |
| | | | C&W L2 | 0.3802 | 0.6198 |
| | | | C&W Linf | 0.7917 | 0.2083 |
| | | | FGSM L1 | 0.7865 | 0.2135 |
| | | | FGSM L2 | 0.6875 | 0.3125 |
| | | | FGSM Linf | 0.2344 | 0.7656 |
| | | | BIM | 0.2083 | 0.7917 |
| | | | MIM | 0.2135 | 0.7865 |
| | | | PGD L1 | 0.7813 | 0.2188 |
| | | | PGD L2 | 0.6094 | 0.3906 |
| | | | PGD Linf | 0.2135 | 0.7865 |

**Table 2.** Performance benchmark results for dataset: German.

| Dataset | ML models | Natural Accuracy | Attack methods | Robust Accuracy | Success Rate |
|---------|-----------|------------------|----------------|-----------------|--------------|
| COMPAS | LR | 0.7933 | DeepFool | 0.2067 | 0.7933 |
| | | | C&W L2 | 0.3054 | 0.6946 |
| | | | C&W Linf | 0.7919 | 0.2081 |
| | | | FGSM L1 | 0.7003 | 0.2997 |
| | | | FGSM L2 | 0.4077 | 0.5923 |
| | | | FGSM Linf | 0.2074 | 0.7926 |
| | | | BIM | 0.2074 | 0.7926 |
| | | | MIM | 0.2074 | 0.7926 |
| | | | PGD L1 | 0.6577 | 0.3423 |
| | | | PGD L2 | 0.4041 | 0.5959 |
| | | | PGD Linf | 0.2074 | 0.7926 |
| | Linear SVC | 0.7976 | DeepFool | 0.2024 | 0.7976 |
| | | | C&W L2 | 0.7919 | 0.2081 |
| | | | C&W Linf | 0.7827 | 0.2173 |
| | | | FGSM L1 | 0.6676 | 0.3324 |
| | | | FGSM L2 | 0.3366 | 0.6634 |
| | | | FGSM Linf | 0.2024 | 0.7976 |
| | | | BIM | 0.2024 | 0.7976 |
| | | | MIM | 0.2024 | 0.7976 |
| | | | PGD L1 | 0.6286 | 0.3714 |
| | | | PGD L2 | 0.3331 | 0.6669 |
| | | | PGD Linf | 0.2024 | 0.7976 |
| | MLP | 0.8089 | DeepFool | 0.1953 | 0.8047 |
| | | | C&W L2 | 0.2791 | 0.7209 |
| | | | C&W Linf | 0.7841 | 0.2159 |
| | | | FGSM L1 | 0.6804 | 0.3196 |
| | | | FGSM L2 | 0.3544 | 0.6456 |
| | | | FGSM Linf | 0.1918 | 0.8082 |
| | | | BIM | 0.1911 | 0.8089 |
| | | | MIM | 0.1911 | 0.8089 |
| | | | PGD L1 | 0.6200 | 0.3800 |
| | | | PGD L2 | 0.3352 | 0.6648 |
| | | | PGD Linf | 0.1911 | 0.8089 |

**Table 3.** Performance benchmark results for dataset: COMPAS.

| Dataset | ML models | Natural Accuracy | Attack methods | Robust Accuracy | Success Rate |
|---------|-----------|------------------|----------------|-----------------|--------------|
| | | | DeepFool | 0.2188 | 0.7578 |
| | | | C&W L2 | 0.2500 | 0.7500 |
| | | | C&W Linf | 0.6875 | 0.3125 |
| | | | LowProFool L2 | 0.2422 | 0.7578 |
| | | | LowProFool Linf | 0.2422 | 0.7578 |
| | | | FGSM L1 | 0.6250 | 0.3750 |
| | LR | 0.7578 | FGSM L2 | 0.4063 | 0.5938 |
| | | | FGSM Linf | 0.2500 | 0.7500 |
| | | | BIM | 0.2500 | 0.7500 |
| | | | MIM | 0.2500 | 0.7500 |
| | | | PGD L1 | 0.6328 | 0.3672 |
| | | | PGD L2 | 0.4063 | 0.5938 |
| | | | PGD Linf | 0.2500 | 0.7500 |
| | | | DeepFool | 0.2422 | 0.7578 |
| | | | C&W L2 | 0.7422 | 0.2578 |
| | | | C&W Linf | 0.7031 | 0.2969 |
| | | | LowProFool L2 | 0.2422 | 0.7578 |
| | | | LowProFool Linf | 0.2422 | 0.7578 |
| | | | FGSM L1 | 0.6250 | 0.3750 |
| Diabetes | Linear SVC | 0.7578 | FGSM L2 | 0.3125 | 0.6875 |
| | | | FGSM Linf | 0.2422 | 0.7578 |
| | | | BIM | 0.2422 | 0.7578 |
| | | | MIM | 0.2422 | 0.7578 |
| | | | PGD L1 | 0.5469 | 0.4531 |
| | | | PGD L2 | 0.3125 | 0.6875 |
| | | | PGD Linf | 0.2422 | 0.7578 |
| | | | DeepFool | 0.2734 | 0.7266 |
| | | | C&W L2 | 0.2813 | 0.7188 |
| | | | C&W Linf | 0.7266 | 0.2734 |
| | | | LowProFool L2 | 0.2734 | 0.7266 |
| | | | LowProFool Linf | 0.2734 | 0.7266 |
| | | | FGSM L1 | 0.6875 | 0.3125 |
| | MLP | 0.7266 | FGSM L2 | 0.4219 | 0.5781 |
| | | | FGSM Linf | 0.2734 | 0.7266 |
| | | | BIM | 0.2734 | 0.7266 |
| | | | MIM | 0.2734 | 0.7266 |
| | | | PGD L1 | 0.6875 | 0.3125 |
| | | | PGD L2 | 0.4141 | 0.5859 |
| | | | PGD Linf | 0.2734 | 0.7266 |

**Table 4.** Performance benchmark results for dataset: Diabetes.

| Dataset | ML models | Natural Accuracy | Attack methods | Robust Accuracy | Success Rate |
|---|---|---|---|---|---|
| | | | DeepFool | 0.0156 | 0.9844 |
| | | | C&W L2 | 0.0938 | 0.9063 |
| | | | C&W Linf | 0.8906 | 0.1094 |
| | | | LowProFool L2 | 0.0156 | 0.9844 |
| | | | LowProFool Linf | 0.0156 | 0.9844 |
| | | | FGSM L1 | 0.9531 | 0.0469 |
| | LR | 0.9844 | FGSM L2 | 0.7656 | 0.2344 |
| | | | FGSM Linf | 0.0313 | 0.9688 |
| | | | BIM | 0.0313 | 0.9688 |
| | | | MIM | 0.0313 | 0.9688 |
| | | | PGD L1 | 0.9531 | 0.0469 |
| | | | PGD L2 | 0.7656 | 0.2344 |
| | | | PGD Linf | 0.0313 | 0.9688 |
| | | | DeepFool | 0.0156 | 0.9844 |
| | | | C&W L2 | 0.9531 | 0.0469 |
| | | | C&W Linf | 0.8750 | 0.1250 |
| | | | LowProFool L2 | 0.0156 | 0.9844 |
| | | | LowProFool Linf | 0.0156 | 0.9844 |
| | | | FGSM L1 | 0.9531 | 0.0469 |
| Breast Cancer | Linear SVC | 0.9844 | FGSM L2 | 0.5156 | 0.4844 |
| | | | FGSM Linf | 0.0156 | 0.9844 |
| | | | BIM | 0.0156 | 0.9844 |
| | | | MIM | 0.0156 | 0.9844 |
| | | | PGD L1 | 0.9531 | 0.0469 |
| | | | PGD L2 | 0.5156 | 0.4844 |
| | | | PGD Linf | 0.0156 | 0.9844 |
| | | | DeepFool | 0.0313 | 0.9688 |
| | | | C&W L2 | 0.1719 | 0.8281 |
| | | | C&W Linf | 0.7969 | 0.2031 |
| | | | LowProFool L2 | 0.0313 | 0.9688 |
| | | | LowProFool Linf | 0.0313 | 0.9688 |
| | | | FGSM L1 | 0.9063 | 0.0938 |
| | MLP | 0.9688 | FGSM L2 | 0.3438 | 0.6563 |
| | | | FGSM Linf | 0.0313 | 0.9688 |
| | | | BIM | 0.0313 | 0.9688 |
| | | | MIM | 0.0156 | 0.9844 |
| | | | PGD L1 | 0.9063 | 0.0938 |
| | | | PGD L2 | 0.7188 | 0.2813 |
| | | | PGD Linf | 0.0313 | 0.9688 |

**Table 5.** Performance benchmark results for dataset: Breast Cancer.

| Dataset | ML models | Attack methods | Spa | L1-norm | L2-norm | Linf-norm | MD | ND | Sens |
|---|---|---|---|---|---|---|---|---|---|
| Adult | LR | DeepFool | 3.9497 | 1.4389 | 0.6355 | 0.4189 | 0.1274 | 0.5444 | 3.0324 |
| | | C&W L2 | 3.7146 | 0.7874 | 0.5217 | 0.4081 | 0.0773 | 0.4190 | 2.5313 |
| | | C&W Linf | 2.1298 | 0.0030 | 0.0020 | 0.0015 | 0.0003 | 0.0016 | 4.6688 |
| | | FGSM L1 | 3.6779 | 0.0762 | 0.0554 | 0.0535 | 0.0046 | 0.0512 | 4.8758 |
| | | FGSM L2 | 3.6779 | 0.3261 | 0.2357 | 0.2271 | 0.0198 | 0.2118 | 5.2120 |
| | | FGSM Linf | 3.6779 | 1.0864 | 0.5644 | 0.3000 | 0.0787 | 0.4560 | 4.0985 |
| | | BIM | 3.6779 | 1.0864 | 0.5644 | 0.3000 | 0.0787 | 0.4560 | 4.0985 |
| | | MIM | 3.6779 | 1.0864 | 0.5644 | 0.3000 | 0.0787 | 0.4560 | 4.0985 |
| | | PGD L1 | 3.6779 | 0.1267 | 0.0918 | 0.0886 | 0.0077 | 0.0834 | 5.0018 |
| | | PGD L2 | 3.6779 | 0.3658 | 0.2632 | 0.2510 | 0.0232 | 0.2307 | 5.3188 |
| | | PGD Linf | 3.6779 | 1.0864 | 0.5644 | 0.3000 | 0.0787 | 0.4560 | 4.0985 |
| | SVM | DeepFool | 3.6946 | 0.1391 | 0.1108 | 0.1086 | 0.0089 | 0.0989 | 5.0922 |
| | | C&W L2 | 2.0699 | 0.0008 | 0.0006 | 0.0005 | 0.0001 | 0.0005 | 4.6686 |
| | | C&W Linf | 2.1289 | 0.0031 | 0.0021 | 0.0016 | 0.0003 | 0.0017 | 4.6671 |
| | | FGSM L1 | 3.6946 | 0.0826 | 0.0661 | 0.0651 | 0.0052 | 0.0600 | 4.9217 |
| | | FGSM L2 | 3.6946 | 0.2977 | 0.2369 | 0.2330 | 0.0187 | 0.2145 | 5.2195 |
| | | FGSM Linf | 3.6946 | 1.0904 | 0.5656 | 0.3000 | 0.0788 | 0.4571 | 4.1120 |
| | | BIM | 3.6946 | 1.0904 | 0.5656 | 0.3000 | 0.0788 | 0.4571 | 4.1120 |
| | | MIM | 3.6946 | 1.0904 | 0.5656 | 0.3000 | 0.0788 | 0.4571 | 4.1120 |
| | | PGD L1 | 3.6946 | 0.1402 | 0.1118 | 0.1100 | 0.0088 | 0.1009 | 5.0738 |
| | | PGD L2 | 3.6946 | 0.3296 | 0.2598 | 0.2533 | 0.0213 | 0.2321 | 5.2993 |
| | | PGD Linf | 3.6946 | 1.0904 | 0.5656 | 0.3000 | 0.0788 | 0.4571 | 4.1120 |
| | MLP | DeepFool | 4.7433 | 3.1593 | 0.9627 | 0.4154 | 0.2861 | 0.7677 | 2.6978 |
| | | C&W L2 | 3.0948 | 0.2609 | 0.1866 | 0.1570 | 0.0277 | 0.1237 | 3.6248 |
| | | C&W Linf | 2.1397 | 0.0033 | 0.0023 | 0.0017 | 0.0004 | 0.0018 | 4.6654 |
| | | FGSM L1 | 3.6890 | 0.0548 | 0.0408 | 0.0393 | 0.0036 | 0.0373 | 4.8234 |
| | | FGSM L2 | 3.6890 | 0.2769 | 0.2030 | 0.1948 | 0.0180 | 0.1788 | 5.2264 |
| | | FGSM Linf | 3.6890 | 1.0907 | 0.5662 | 0.2999 | 0.0779 | 0.4511 | 4.1864 |
| | | BIM | 3.6700 | 0.8957 | 0.4956 | 0.2997 | 0.0633 | 0.4036 | 4.9685 |
| | | MIM | 3.6920 | 1.0915 | 0.5663 | 0.3000 | 0.0786 | 0.4548 | 4.1863 |
| | | PGD L1 | 3.6906 | 0.1426 | 0.1050 | 0.1010 | 0.0092 | 0.0930 | 5.0502 |
| | | PGD L2 | 3.6945 | 0.3395 | 0.2533 | 0.2437 | 0.0225 | 0.2242 | 5.3342 |
| | | PGD Linf | 3.6920 | 1.0915 | 0.5663 | 0.3000 | 0.0786 | 0.4548 | 4.1863 |

**Table 6.** Imperceptibility benchmark results for dataset: Adult.

| Dataset | ML models | Attack methods | Spa | L1-norm | L2-norm | Linf-norm | MD | ND | Sens |
|---|---|---|---|---|---|---|---|---|---|
| German | LR | DeepFool | 4.6458 | 1.5837 | 0.6435 | 0.3364 | 0.2109 | 0.6188 | 2.7336 |
| | | C&W L2 | 5.0104 | 1.1802 | 0.6234 | 0.4146 | 0.1540 | 0.6211 | 2.6634 |
| | | C&W Linf | 3.7552 | 0.0044 | 0.0023 | 0.0014 | 0.0006 | 0.0023 | 2.9240 |
| | | FGSM L1 | 4.3073 | 0.0617 | 0.0327 | 0.0205 | 0.0070 | 0.0327 | 2.9551 |
| | | FGSM L2 | 4.3073 | 0.3588 | 0.1907 | 0.1206 | 0.0409 | 0.1907 | 3.0578 |
| | | FGSM Linf | 4.3073 | 1.1882 | 0.5837 | 0.3000 | 0.1519 | 0.5830 | 2.9858 |
| | | BIM | 4.3073 | 1.1882 | 0.5837 | 0.3000 | 0.1519 | 0.5830 | 2.9858 |
| | | MIM | 4.3073 | 1.1882 | 0.5837 | 0.3000 | 0.1519 | 0.5830 | 2.9858 |
| | | PGD L1 | 4.3073 | 0.0988 | 0.0525 | 0.0330 | 0.0112 | 0.0525 | 2.9747 |
| | | PGD L2 | 4.3073 | 0.4197 | 0.2245 | 0.1439 | 0.0484 | 0.2245 | 3.0913 |
| | | PGD Linf | 4.3073 | 1.1882 | 0.5837 | 0.3000 | 0.1519 | 0.5830 | 2.9858 |
| | SVM | DeepFool | 4.3125 | 0.7560 | 0.4128 | 0.3027 | 0.0859 | 0.4128 | 3.0057 |
| | | C&W L2 | 3.6927 | 0.0002 | 0.0001 | 0.0001 | 0.0000 | 0.0001 | 2.9242 |
| | | C&W Linf | 3.7500 | 0.0031 | 0.0016 | 0.0010 | 0.0004 | 0.0016 | 2.9243 |
| | | FGSM L1 | 4.3125 | 0.0635 | 0.0342 | 0.0243 | 0.0071 | 0.0342 | 2.9562 |
| | | FGSM L2 | 4.3125 | 0.3562 | 0.1926 | 0.1369 | 0.0399 | 0.1926 | 3.0590 |
| | | FGSM Linf | 4.3125 | 1.1894 | 0.5840 | 0.3000 | 0.1521 | 0.5832 | 2.9933 |
| | | BIM | 4.3125 | 1.1894 | 0.5840 | 0.3000 | 0.1521 | 0.5832 | 2.9933 |
| | | MIM | 4.3125 | 1.1894 | 0.5840 | 0.3000 | 0.1521 | 0.5832 | 2.9933 |
| | | PGD L1 | 4.3125 | 0.1065 | 0.0577 | 0.0411 | 0.0119 | 0.0577 | 2.9798 |
| | | PGD L2 | 4.3125 | 0.4241 | 0.2307 | 0.1653 | 0.0481 | 0.2307 | 3.0948 |
| | | PGD Linf | 4.3125 | 1.1894 | 0.5840 | 0.3000 | 0.1521 | 0.5832 | 2.9933 |
| | MLP | DeepFool | 6.0000 | 4.0822 | 1.1269 | 0.4170 | 0.4722 | 1.0040 | 2.6485 |
| | | C&W L2 | 4.7344 | 0.7037 | 0.4396 | 0.3569 | 0.0999 | 0.4393 | 2.6641 |
| | | C&W Linf | 3.7448 | 0.0026 | 0.0014 | 0.0009 | 0.0003 | 0.0014 | 2.9244 |
| | | FGSM L1 | 4.4219 | 0.0376 | 0.0211 | 0.0157 | 0.0046 | 0.0211 | 2.9436 |
| | | FGSM L2 | 4.4219 | 0.2354 | 0.1323 | 0.0991 | 0.0289 | 0.1323 | 3.0183 |
| | | FGSM Linf | 4.4219 | 1.2242 | 0.5932 | 0.3000 | 0.1654 | 0.5928 | 3.1883 |
| | | BIM | 4.2188 | 1.0332 | 0.5301 | 0.2994 | 0.1415 | 0.5297 | 2.8342 |
| | | MIM | 4.3281 | 1.1869 | 0.5826 | 0.3000 | 0.1595 | 0.5822 | 2.9970 |
| | | PGD L1 | 4.4167 | 0.0708 | 0.0398 | 0.0296 | 0.0087 | 0.0398 | 2.9605 |
| | | PGD L2 | 4.4063 | 0.3088 | 0.1744 | 0.1317 | 0.0381 | 0.1744 | 3.0418 |
| | | PGD Linf | 4.3281 | 1.1869 | 0.5826 | 0.3000 | 0.1595 | 0.5822 | 2.9970 |

**Table 7.** Imperceptibility benchmark results for dataset: Adult.

| Dataset | ML models | Attack methods | Spa | L1-norm | L2-norm | Linf-norm | MD | ND | Sens |
|---|---|---|---|---|---|---|---|---|---|
| COMPAS | LR | DeepFool | 3.9382 | 0.6875 | 0.4108 | 0.3257 | 0.0513 | 0.2679 | 4.0863 |
| | | C&W L2 | 3.8040 | 0.3211 | 0.2426 | 0.2147 | 0.0270 | 0.1392 | 5.7085 |
| | | C&W Linf | 3.2202 | 0.0033 | 0.0021 | 0.0016 | 0.0003 | 0.0018 | 7.4214 |
| | | FGSM L1 | 3.9311 | 0.2116 | 0.1286 | 0.1032 | 0.0160 | 0.0619 | 8.4434 |
| | | FGSM L2 | 3.9311 | 0.4495 | 0.2773 | 0.2271 | 0.0345 | 0.1322 | 7.4899 |
| | | FGSM Linf | 3.9311 | 0.9953 | 0.5289 | 0.3000 | 0.0576 | 0.3357 | 6.1583 |
| | | BIM | 3.9311 | 0.9953 | 0.5289 | 0.3000 | 0.0576 | 0.3357 | 6.1583 |
| | | MIM | 3.9311 | 0.9953 | 0.5289 | 0.3000 | 0.0576 | 0.3357 | 6.1583 |
| | | PGD L1 | 3.9311 | 0.2470 | 0.1524 | 0.1243 | 0.0190 | 0.0727 | 8.4531 |
| | | PGD L2 | 3.9311 | 0.4696 | 0.2961 | 0.2478 | 0.0371 | 0.1440 | 7.2799 |
| | | PGD Linf | 3.9311 | 0.9953 | 0.5289 | 0.3000 | 0.0576 | 0.3357 | 6.1583 |
| | SVM | DeepFool | 3.9361 | 0.3910 | 0.2360 | 0.1833 | 0.0290 | 0.1178 | 8.1577 |
| | | C&W L2 | 3.2003 | 0.0024 | 0.0017 | 0.0015 | 0.0002 | 0.0013 | 7.4227 |
| | | C&W Linf | 3.2131 | 0.0033 | 0.0021 | 0.0016 | 0.0003 | 0.0017 | 7.4202 |
| | | FGSM L1 | 3.9361 | 0.2156 | 0.1301 | 0.1006 | 0.0160 | 0.0642 | 8.4932 |
| | | FGSM L2 | 3.9361 | 0.4534 | 0.2779 | 0.2199 | 0.0342 | 0.1359 | 7.5689 |
| | | FGSM Linf | 3.9361 | 1.0023 | 0.5314 | 0.3000 | 0.0577 | 0.3355 | 6.1877 |
| | | BIM | 3.9361 | 1.0023 | 0.5314 | 0.3000 | 0.0577 | 0.3355 | 6.1877 |
| | | MIM | 3.9361 | 1.0023 | 0.5314 | 0.3000 | 0.0577 | 0.3355 | 6.1877 |
| | | PGD L1 | 3.9361 | 0.2507 | 0.1536 | 0.1208 | 0.0190 | 0.0747 | 8.5202 |
| | | PGD L2 | 3.9361 | 0.4733 | 0.2963 | 0.2403 | 0.0367 | 0.1466 | 7.3665 |
| | | PGD Linf | 3.9361 | 1.0023 | 0.5314 | 0.3000 | 0.0577 | 0.3355 | 6.1877 |
| | MLP | DeepFool | 3.9929 | 0.8662 | 0.4637 | 0.3198 | 0.0856 | 0.2947 | 3.6962 |
| | | C&W L2 | 3.8111 | 0.3297 | 0.2459 | 0.2137 | 0.0277 | 0.1461 | 5.4770 |
| | | C&W Linf | 3.2102 | 0.0031 | 0.0020 | 0.0015 | 0.0003 | 0.0017 | 7.4239 |
| | | FGSM L1 | 3.9197 | 0.1671 | 0.0990 | 0.0761 | 0.0118 | 0.0523 | 8.2735 |
| | | FGSM L2 | 3.9197 | 0.4437 | 0.2668 | 0.2077 | 0.0319 | 0.1331 | 7.6347 |
| | | FGSM Linf | 3.9197 | 0.9591 | 0.5187 | 0.3000 | 0.0568 | 0.3247 | 6.0415 |
| | | BIM | 3.8402 | 0.8141 | 0.4717 | 0.3000 | 0.0536 | 0.2874 | 5.8091 |
| | | MIM | 3.9006 | 0.9337 | 0.5109 | 0.3000 | 0.0563 | 0.3182 | 5.9662 |
| | | PGD L1 | 3.9190 | 0.2096 | 0.1263 | 0.0982 | 0.0152 | 0.0631 | 8.3842 |
| | | PGD L2 | 3.9134 | 0.4659 | 0.2871 | 0.2286 | 0.0349 | 0.1435 | 7.4439 |
| | | PGD Linf | 3.9006 | 0.9337 | 0.5109 | 0.3000 | 0.0563 | 0.3182 | 5.9662 |

**Table 8.** Imperceptibility benchmark results for dataset: COMPAS.

| Dataset | ML models | Attack methods | Spa | L1-norm | L2-norm | Linf-norm | MD | ND | Sens |
|---|---|---|---|---|---|---|---|---|---|
| Diabetes | LR | DeepFool | 7.4375 | 0.5727 | 0.2719 | 0.1938 | 0.0500 | 0.1697 | 3.7474 |
| | | C&W L2 | 8.0000 | 0.3429 | 0.1942 | 0.1491 | 0.0326 | 0.1451 | 3.7797 |
| | | C&W Linf | 6.9609 | 0.0319 | 0.0129 | 0.0061 | 0.0022 | 0.0129 | 4.0271 |
| | | LowProFool L2 | 7.4375 | 1.6105 | 0.7545 | 0.5228 | 0.1365 | 0.4429 | 3.0605 |
| | | LowProFool Linf | 7.4375 | 1.6171 | 0.7487 | 0.5067 | 0.1354 | 0.4406 | 3.0552 |
| | | FGSM L1 | 7.4375 | 0.2951 | 0.1421 | 0.1041 | 0.0258 | 0.1331 | 4.0163 |
| | | FGSM L2 | 7.4375 | 0.6202 | 0.2988 | 0.2190 | 0.0543 | 0.2140 | 3.8414 |
| | | FGSM Linf | 7.4375 | 2.1037 | 0.7842 | 0.3000 | 0.1375 | 0.4924 | 3.3679 |
| | | BIM | 7.4375 | 2.1037 | 0.7842 | 0.3000 | 0.1375 | 0.4924 | 3.3679 |
| | | MIM | 7.4375 | 2.1037 | 0.7842 | 0.3000 | 0.1375 | 0.4924 | 3.3679 |
| | | PGD L1 | 7.4375 | 0.3000 | 0.1445 | 0.1059 | 0.0263 | 0.1350 | 4.0171 |
| | | PGD L2 | 7.4375 | 0.6232 | 0.3000 | 0.2191 | 0.0544 | 0.2145 | 3.8506 |
| | | PGD Linf | 7.4375 | 2.1037 | 0.7842 | 0.3000 | 0.1375 | 0.4924 | 3.3679 |
| | SVM | DeepFool | 7.4609 | 0.4818 | 0.2305 | 0.1592 | 0.0384 | 0.1738 | 3.9778 |
| | | C&W L2 | 6.8438 | 0.0168 | 0.0091 | 0.0066 | 0.0014 | 0.0091 | 4.0361 |
| | | C&W Linf | 6.9297 | 0.0455 | 0.0178 | 0.0080 | 0.0029 | 0.0178 | 4.0331 |
| | | LowProFool L2 | 7.4609 | 1.3735 | 0.6557 | 0.4548 | 0.1081 | 0.3950 | 3.2364 |
| | | LowProFool Linf | 7.4609 | 1.3735 | 0.6557 | 0.4548 | 0.1081 | 0.3950 | 3.2364 |
| | | FGSM L1 | 7.4609 | 0.2919 | 0.1390 | 0.0959 | 0.0231 | 0.1306 | 4.0607 |
| | | FGSM L2 | 7.4609 | 0.6257 | 0.2983 | 0.2061 | 0.0495 | 0.2164 | 3.9135 |
| | | FGSM Linf | 7.4609 | 2.0919 | 0.7808 | 0.3000 | 0.1342 | 0.4783 | 3.3899 |
| | | BIM | 7.4609 | 2.0919 | 0.7808 | 0.3000 | 0.1342 | 0.4783 | 3.3899 |
| | | MIM | 7.4609 | 2.0919 | 0.7808 | 0.3000 | 0.1342 | 0.4783 | 3.3899 |
| | | PGD L1 | 7.4609 | 0.3000 | 0.1430 | 0.0987 | 0.0238 | 0.1336 | 4.0596 |
| | | PGD L2 | 7.4609 | 0.6286 | 0.3000 | 0.2074 | 0.0497 | 0.2169 | 3.9174 |
| | | PGD Linf | 7.4609 | 2.0919 | 0.7808 | 0.3000 | 0.1342 | 0.4783 | 3.3899 |
| | MLP | DeepFool | 7.5391 | 0.7205 | 0.2912 | 0.1701 | 0.0560 | 0.1873 | 4.0465 |
| | | C&W L2 | 8.0000 | 0.4421 | 0.2114 | 0.1393 | 0.0344 | 0.1516 | 3.9052 |
| | | C&W Linf | 6.9922 | 0.0474 | 0.0188 | 0.0087 | 0.0033 | 0.0188 | 4.0394 |
| | | LowProFool L2 | 7.5547 | 1.4901 | 0.6017 | 0.3471 | 0.1128 | 0.3581 | 3.7346 |
| | | LowProFool Linf | 7.5547 | 1.5197 | 0.6108 | 0.3474 | 0.1147 | 0.3632 | 3.7122 |
| | | FGSM L1 | 7.5391 | 0.2879 | 0.1168 | 0.0683 | 0.0219 | 0.1130 | 4.1771 |
| | | FGSM L2 | 7.5391 | 0.7234 | 0.2936 | 0.1714 | 0.0550 | 0.2162 | 4.1953 |
| | | FGSM Linf | 7.5391 | 2.1600 | 0.7939 | 0.3000 | 0.1463 | 0.4451 | 3.4685 |
| | | BIM | 7.8203 | 2.1187 | 0.7765 | 0.3000 | 0.1458 | 0.4290 | 3.7815 |
| | | MIM | 7.6250 | 2.1889 | 0.7999 | 0.3000 | 0.1492 | 0.4519 | 3.5542 |
| | | PGD L1 | 7.5469 | 0.3000 | 0.1221 | 0.0714 | 0.0229 | 0.1174 | 4.1876 |
| | | PGD L2 | 7.5547 | 0.7401 | 0.3000 | 0.1749 | 0.0562 | 0.2179 | 4.2087 |
| | | PGD Linf | 7.6250 | 2.1889 | 0.7999 | 0.3000 | 0.1492 | 0.4519 | 3.5542 |

**Table 9.** Imperceptibility benchmark results for dataset: Diabetes.

| Dataset | ML models | Attack methods | Spa | L1-norm | L2-norm | Linf-norm | MD | ND | Sens |
|---|---|---|---|---|---|---|---|---|---|
| Breast Cancer | LR | DeepFool | 29.9688 | 8.0252 | 1.7005 | 0.5240 | 0.8427 | 0.7984 | 3.1569 |
| | | C&W L2 | 29.8125 | 1.7654 | 0.4392 | 0.1854 | 0.2088 | 0.3032 | 5.8483 |
| | | C&W Linf | 29.7031 | 0.0743 | 0.0144 | 0.0036 | 0.0068 | 0.0144 | 5.8651 |
| | | LowProFool L2 | 29.9688 | 5.9440 | 1.3016 | 0.4537 | 0.6694 | 0.4952 | 4.6033 |
| | | LowProFool Linf | 29.9688 | 6.2122 | 1.3525 | 0.4515 | 0.6921 | 0.5138 | 4.4905 |
| | | FGSM L1 | 29.9688 | 0.2999 | 0.0652 | 0.0227 | 0.0334 | 0.0652 | 6.0003 |
| | | FGSM L2 | 29.9688 | 1.3793 | 0.2999 | 0.1044 | 0.1537 | 0.2776 | 6.3440 |
| | | FGSM Linf | 29.9688 | 8.0973 | 1.5263 | 0.3000 | 0.7075 | 0.8777 | 4.2016 |
| | | BIM | 29.9688 | 8.0973 | 1.5263 | 0.3000 | 0.7075 | 0.8777 | 4.2016 |
| | | MIM | 29.9688 | 8.0973 | 1.5263 | 0.3000 | 0.7075 | 0.8777 | 4.2016 |
| | | PGD L1 | 29.9688 | 0.3000 | 0.0652 | 0.0227 | 0.0334 | 0.0652 | 6.0004 |
| | | PGD L2 | 29.9688 | 1.3789 | 0.3000 | 0.1045 | 0.1537 | 0.2776 | 6.3464 |
| | | PGD Linf | 29.9688 | 8.0973 | 1.5263 | 0.3000 | 0.7075 | 0.8777 | 4.2016 |
| | SVM | DeepFool | 29.9375 | 1.8211 | 0.3923 | 0.1408 | 0.1540 | 0.3365 | 6.5070 |
| | | C&W L2 | 29.7031 | 0.0072 | 0.0018 | 0.0009 | 0.0006 | 0.0018 | 5.8490 |
| | | C&W Linf | 29.7031 | 0.0901 | 0.0175 | 0.0044 | 0.0063 | 0.0175 | 5.8615 |
| | | LowProFool L2 | 29.9375 | 4.3552 | 0.9481 | 0.3485 | 0.3789 | 0.6507 | 5.4292 |
| | | LowProFool Linf | 29.9375 | 4.3462 | 0.9452 | 0.3457 | 0.3773 | 0.6496 | 5.4276 |
| | | FGSM L1 | 29.9375 | 0.2998 | 0.0645 | 0.0228 | 0.0251 | 0.0645 | 5.9785 |
| | | FGSM L2 | 29.9375 | 1.3879 | 0.2991 | 0.1061 | 0.1168 | 0.2899 | 6.2360 |
| | | FGSM Linf | 29.9375 | 7.8941 | 1.4978 | 0.3000 | 0.5639 | 0.9559 | 4.0904 |
| | | BIM | 29.9375 | 7.8941 | 1.4978 | 0.3000 | 0.5639 | 0.9559 | 4.0904 |
| | | MIM | 29.9375 | 7.8941 | 1.4978 | 0.3000 | 0.5639 | 0.9559 | 4.0904 |
| | | PGD L1 | 29.9375 | 0.3000 | 0.0646 | 0.0228 | 0.0252 | 0.0646 | 5.9787 |
| | | PGD L2 | 29.9375 | 1.3904 | 0.3000 | 0.1068 | 0.1174 | 0.2901 | 6.2489 |
| | | PGD Linf | 29.9375 | 7.8941 | 1.4978 | 0.3000 | 0.5639 | 0.9559 | 4.0904 |
| | MLP | DeepFool | 29.8438 | 6.8163 | 1.5350 | 0.5677 | 0.4171 | 1.2586 | 3.2212 |
| | | C&W L2 | 29.8125 | 1.0338 | 0.2793 | 0.1502 | 0.0623 | 0.2611 | 5.5206 |
| | | C&W Linf | 29.7031 | 0.1303 | 0.0255 | 0.0067 | 0.0036 | 0.0255 | 5.8420 |
| | | LowProFool L2 | 29.9531 | 7.3195 | 1.7114 | 0.7110 | 0.4691 | 1.4113 | 3.2434 |
| | | LowProFool Linf | 29.9375 | 7.3255 | 1.6950 | 0.7055 | 0.4544 | 1.4083 | 3.2709 |
| | | FGSM L1 | 29.8438 | 0.2982 | 0.0677 | 0.0280 | 0.0154 | 0.0677 | 5.9247 |
| | | FGSM L2 | 29.8438 | 1.2835 | 0.2915 | 0.1231 | 0.0677 | 0.2909 | 6.0041 |
| | | FGSM Linf | 29.8438 | 7.3804 | 1.4229 | 0.3000 | 0.2807 | 1.2486 | 3.9272 |
| | | BIM | 29.9531 | 7.7634 | 1.4788 | 0.3000 | 0.4513 | 1.1551 | 4.1319 |
| | | MIM | 29.8594 | 7.6802 | 1.4672 | 0.3000 | 0.3552 | 1.2402 | 3.7860 |
| | | PGD L1 | 29.8438 | 0.3000 | 0.0683 | 0.0280 | 0.0158 | 0.0683 | 5.9286 |
| | | PGD L2 | 29.8438 | 1.3112 | 0.3000 | 0.1225 | 0.0732 | 0.2993 | 6.0677 |
| | | PGD Linf | 29.8594 | 7.6802 | 1.4672 | 0.3000 | 0.3552 | 1.2402 | 3.7860 |

**Table 10.** Imperceptibility benchmark results for dataset: Breast Cancer.