

Xss

Dataset: plain\_xss

Lstm model in the original paper

Running on a personal computer 4060 (note the loaded dataset xss)

```
Model loaded successfully.
Loading test data from: data/xss_dataset_finaltest_X and data/xss_dataset_finaltest_Y
Test data loaded successfully.
8277 samples in the final test set.
Percentage of vulnerable samples: 8.66%
Absolute number of vulnerable samples: 717
Starting predictions...

Evaluation Metrics:
Accuracy: 0.9136
Precision: 0.5008
Recall: 0.8480
F1 Score: 0.6297
```

CNN ResNet

Running on colab A100 (note the loaded dataset xss)

```
Model loaded successfully.
Loading test data from: data/xss_dataset_finaltest_X and data/xss_dataset_finaltest_Y
Test data loaded successfully.
8277 samples in the final test set.
Percentage of vulnerable samples: 9.00%
Absolute number of vulnerable samples: 745
Starting predictions...
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
I0000 00:00:1732543930.981629 15315 service.cc:146] XLA service 0x79b9ac063160 initialized for platform CUDA (this does
2024-11-25 14:12:11.051543: I tensorflow/compiler/mlir/tensorflow/utils/dump_mlir_util.cc:268] disabling MLIR crash repr
2024-11-25 14:12:11.287271: I external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:531] Loaded cuDNN version 8906
I0000 00:00:1732543932.721800 15315 device_compiler.h:188] Compiled cluster using XLA! This line is logged at most once

Evaluation Metrics:
Accuracy: 0.9756
Precision: 0.9563
Recall: 0.7638
F1 Score: 0.8493
```

CNN ResNet training parameters (note the loaded dataset xss)

```
Creating training dataset... (xss)
Creating validation dataset...
Creating finaltest dataset...
Train length: 38625
Test length: 8277
Finaltesting length: 8277
time: 13:48
X_train shape: (38625, 200, 200), dtype: float32
y_train shape: (38625,), dtype: int64
X_test shape: (8277, 200, 200), dtype: float32
y_test shape: (8277,), dtype: int64
X_finaltest shape: (8277, 200, 200), dtype: float32
y_finaltest shape: (8277,), dtype: int64
numpy array done. 13:48
38625 samples in the training set.
8277 samples in the validation set.
8277 samples in the final test set.
percentage of vulnerable samples: 8.62%
absolute amount of vulnerable samples in test set: 770
Starting CNNResNet: 13:48
Dropout: 0.2
Neurons: 100
Optimizer: adam
Epochs: 20
Batch Size: 128
max length: 200
```

remote\_code\_execution

Dataset: plain\_remote\_code\_execution

Lstm model in the original paper

Running on a personal computer 4060 (note the loaded dataset

remote\_code\_execution)

```
Model loaded successfully.
Loading test data from: data/remote_code_execution_dataset_finaltest_X and data/remote_code_execution_dataset_finaltest_Y
Test data loaded successfully.
14412 samples in the final test set.
Percentage of vulnerable samples: 9.04%
Absolute number of vulnerable samples: 1303
Starting predictions...

Evaluation Metrics:
Accuracy: 0.9584
Precision: 0.8308
Recall: 0.6784
F1 Score: 0.7469
```

CNN ResNet

Running on colab A100 (note the loaded dataset remote\_code\_execution)

```
Model loaded successfully.
Loading test data from: data/remote_code_execution_dataset_finaltest_X and data/remote_code_execution_dataset_finaltest_Y
Test data loaded successfully.
14412 samples in the final test set.
Percentage of vulnerable samples: 8.89%
Absolute number of vulnerable samples: 1281
Starting predictions...
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
I0000 00:00:1732543983.732261 15656 service.cc:146] XLA service 0x788878011a40 initialized for platform CUDA (this does not
I0000 00:00:1732543983.732298 15656 service.cc:154] StreamExecutor device (0): NVIDIA A100-SXM4-40GB, Compute Capability 8.0
2024-11-25 14:13:03.749825: I tensorflow/compiler/mlir/tensorflow/utils/dump_mlir_util.cc:268] disabling MLIR crash reproducer
2024-11-25 14:13:03.867610: I external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:531] Loaded cuDNN version 8906
I0000 00:00:1732543984.663532 15656 device_compiler.h:188] Compiled cluster using XLA! This line is logged at most once f

Evaluation Metrics:
Accuracy: 0.9791
Precision: 0.9291
Recall: 0.8283
F1 Score: 0.8758
```

CNN ResNet training parameters (note the loaded dataset

remote\_code\_execution)

```
Creating training dataset... (remote_code_execution)
Creating validation dataset...
Creating finaltest dataset...
Train length: 67255
Test length: 14411
Finaltesting length: 14412
time: 13:39
X_train shape: (67255, 200, 200), dtype: float32
y_train shape: (67255,), dtype: int64
X_test shape: (14411, 200, 200), dtype: float32
y_test shape: (14411,), dtype: int64
X_finaltest shape: (14412, 200, 200), dtype: float32
y_finaltest shape: (14412,), dtype: int64
numpy array done. 13:40
67255 samples in the training set.
14411 samples in the validation set.
14412 samples in the final test set.
percentage of vulnerable samples: 8.99%
absolute amount of vulnerable samples in test set: 1254
Starting CNNResNet: 13:40
Dropout: 0.2
Neurons: 100
Optimizer: adam
Epochs: 20
Batch Size: 128
max length: 200
```

path\_disclosure

Dataset: plain\_path\_disclosure

Lstm model in the original paper

Running on a personal computer 4060 (note the loaded dataset

path\_disclosure)

```
Model loaded successfully.
Loading test data from: data/path_disclosure_dataset_finaltest_X and data/path_disclosure_dataset_finaltest_Y
Test data loaded successfully.
19680 samples in the final test set.
Percentage of vulnerable samples: 11.76%
Absolute number of vulnerable samples: 2315
Starting predictions...

Evaluation Metrics:
Accuracy: 0.8892
Precision: 0.6810
Recall: 0.1097
F1 Score: 0.1890
```

CNN ResNet

Running on colab A100 (note the loaded dataset path\_disclosure)

```
Model loaded successfully.
Loading test data from: data/path_disclosure_dataset_finaltest_X and data/path_disclosure_dataset_finaltest_Y
Test data loaded successfully.
19680 samples in the final test set.
Percentage of vulnerable samples: 11.73%
Absolute number of vulnerable samples: 2308
Starting predictions...
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
I0000 00:00:1732544770.205036 19923 service.cc:146] XLA service 0x7cd6380511d0 initialized for platform CUDA (
I0000 00:00:1732544770.205073 19923 service.cc:154] StreamExecutor device (0): NVIDIA A100-SXM4-40GB, Comput
2024-11-25 14:26:10.272777: I tensorflow/compiler/mlir/tensorflow/utils/dump_mlir_util.cc:268] disabling MLIR cr
2024-11-25 14:26:10.506586: I external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:531] Loaded cuDNN version
I0000 00:00:1732544771.946794 19923 device_compiler.h:188] Compiled cluster using XLA! This line is logged at

Evaluation Metrics:
Accuracy: 0.9686
Precision: 0.9159
Recall: 0.8068
F1 Score: 0.8579
```

CNN ResNet training parameters (note the loaded dataset path\_disclosure)

```
Creating training dataset... (path_disclosure)
Creating validation dataset...
Creating finaltest dataset...
Train length: 91842
Test length: 19681
Finaltesting length: 19680
time: 14:19
X_train shape: (91842, 200, 200), dtype: float32
y_train shape: (91842,), dtype: int64
X_test shape: (19681, 200, 200), dtype: float32
y_test shape: (19681,), dtype: int64
X_finaltest shape: (19680, 200, 200), dtype: float32
y_finaltest shape: (19680,), dtype: int64
numpy array done. 14:20
91842 samples in the training set.
19681 samples in the validation set.
19680 samples in the final test set.
percentage of vulnerable samples: 11.68%
absolute amount of vulnerable samples in test set: 2235
Starting CNNResNet: 14:20
Dropout: 0.2
Neurons: 100
Optimizer: adam
Epochs: 20
Batch Size: 128
max length: 200
```