CSC 180-01 Intelligent Systems (Fall 2024)

Title: - Modern Low Footprint Cyber Attack Detection

Due at 10:30 am- Wednesday, October 09, 2024

Name	Student ID
Taekjin Jung	303293432
Illya Gordyy	302682939
Jenil Shingala	302796429
Danny Phan	301698774

1. Problem Statement

This project aims to build a network intrusion detector capable of distinguishing between bad connections (intrusions or attacks) and good normal connections. The problem is modeled as a binary classification task using the UNSW-NB15 dataset, which reflects modern low footprint attacks. The goal is to compare the performance of Fully-Connected Neural Networks (FCNNs) and Convolutional Neural Networks (CNNs) in detecting network intrusions.

2. Methodology

Data Preparation:

- Used a subset of the UNSW-NB15 dataset: UNSW_NB15_training-set.csv (175,341 records) and UNSW_NB15_testing-set.csv (82,332 records).
- Removed records with categorical values that only appear in either training or test data.
- Dropped rows with missing values.
- Encoded categorical features and normalized numeric features.

Model Development:

- Implemented two types of neural networks: a) Fully-Connected Neural Networks (FCNNs) b) Convolutional Neural Networks (CNNs)
- Used EarlyStopping and ModelCheckpoint during training.

Tuned hyperparameters:

- Activation functions: ReLU, Sigmoid, Tanh +) Leaky ReLu
- Number of layers and neuron counts
- Optimizers: Adam and SGD
- Kernel number and kernel size (for CNN only)

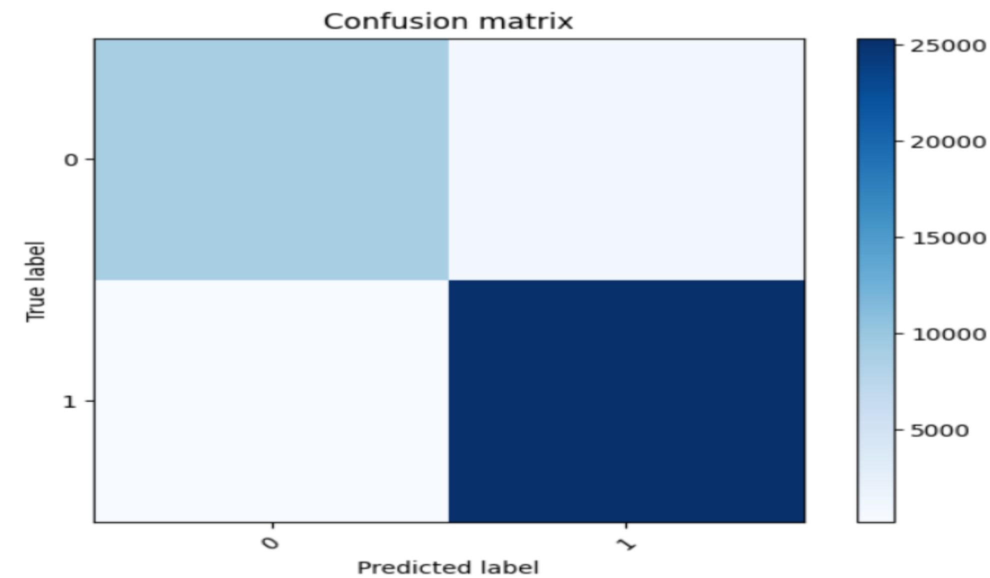
Evaluation Metrics:

- Recall, Precision, and F1-score for both attacks and normal connections
- Confusion matrix
- ROC curve for the best model (in terms of F1-score for intrusion)

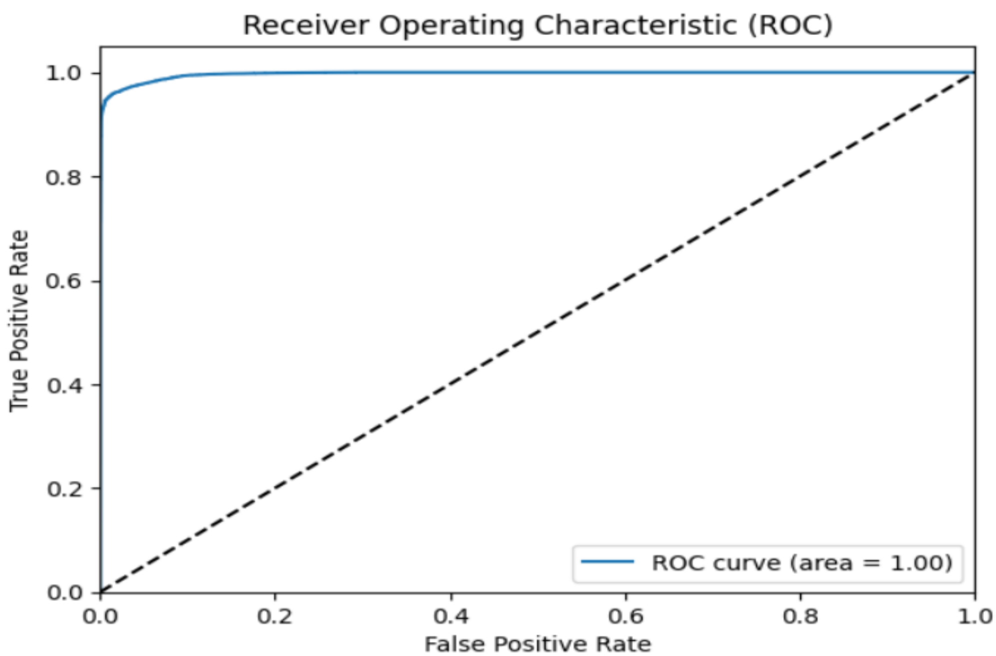
3. Experimental Results and Analysis

Hyperparameters:

Model	Activation	Optimizer	Kernal Number	Neuron count	Kernal size	Accuracy	Avg F1
1	reLu	Adam	64-128	128-2	(3,1)	0.9589	0.9580
2	Tanh	Adam	64-128	128-2	(3,1)	0.9698	0.9695
3	reLu	Adam	128-64	128-2	(3,1)	0.9391	0.9367
4	Leaky ReLu	Adam	64-128	128-2	(3,1)	0.9500	0.9484
5	reLu	Adam	64-128	128-2	(4,1)	0.9555	0.9543
6	reLu	Adam	64-128	256-2	(3,1)	0.9376	0.9349
7	Tanh	SGD	64-128	128-2	(3,1)	0.9446	0.9426



	precision	recall	f1-score	support
0	0.97	0.91	0.94	9624
1	0.97	0.99	0.98	25554
accuracy			0.97	35178
macro avg	0.97	0.95	0.96	35178
weighted avg	0.97	0.97	0.97	35178



As a result, we got the best F1 Score : 0.9698 with a model trained by Tanh (Adam), 64-128 (number of Kernals), and Adam (optimizer), 128-2 (number of neurons). The second figure is the lift chart with the Roc cover.

4. Task Division and Project Reflection

Name	Task
Taekjin Jung	Data encoding and splitting, FCNN, Additional Feature
Illya Gordy	CNN, Training/Testing model, Visualization
Jenil Shingala	Data management, testing different hyperparameters, report
Danny Phan	FCNN, debugging

Challenges:

This project provided valuable insights into the application of deep learning techniques for cybersecurity. Key learnings include:

1. The importance of careful data preprocessing, especially when dealing with categorical features in separate train and test sets.
2. Shape checking – matching the dimension and x, y values between training sets and filtering non-common data in two comparative data sets
3. The effectiveness of CNNs in capturing complex patterns in network traffic data, connecting FCNN to match output counts and neuron counts using ‘Softmax’ to allocate.
4. The critical role of hyperparameter tuning in optimizing model performance.
5. Data cleaning

Learning Outcome:

- Work with big datasets and clean up messy data
- Build and use two types of machine learning models: Fully-Connected Neural Networks and Convolutional Neural Networks
- Different types of layers of CNN
- Adjust different parts of our machine learning models to make them work better
- Work as a team on a big data project

- Think about the right and wrong ways of using machine learning for security
- The importance of matching shape and dimension in input/output values