**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG**



**BÁO CÁO TIẾN ĐỘ**

**Intrusion Detection System based on Federated Learning**

**Thành viên :**

1. Đậu Trọng Hoàng MSSV: 19521529

2. Lê Chí Đại MSSV: 19520441

**Giảng viên hướng dẫn :** **PGS. TS. Lê Trung Quân**

**MỤC LỤC NỘI DUNG**

[I. Nghiên cứu cơ sở lý thuyết 6](#_Toc121394340)

[A. Mạng nơ-ron – Neural Network 6](#_Toc121394341)

[B. Mạng nơ-ron tích chập – Convolutional Neural Network 6](#_Toc121394342)

[C. Tâp dữ liệu không đều – Imbalanced Dataset 7](#_Toc121394343)

[D. SMOTE – Synthetic Minority Over-sampling 7](#_Toc121394344)

[E. ENN – Edited Nearest Neighbors 7](#_Toc121394345)

[F. Máy học cộng tác – Federated Learning 7](#_Toc121394346)

[II. Khảo sát các nghiên cứu liên quan 8](#_Toc121394347)

[A. Model cho mạng nơ-ron tích chập 8](#_Toc121394348)

[B. Tập dữ liệu 8](#_Toc121394349)

[C. Tiền xử lý dữ liệu 9](#_Toc121394350)

[D. Phương pháp đánh giá hiệu quả 10](#_Toc121394351)

[III. Thực nghiệm 11](#_Toc121394352)

[A. Trên mạng nơ-ron tích chập sâu 11](#_Toc121394353)

[1. Local model 11](#_Toc121394354)

[B. Cải thiện với SMOTE-ENN 11](#_Toc121394355)

[1. Quy trình cải thiện tập dữ liệu 11](#_Toc121394356)

**MỤC LỤC HÌNH ẢNH**

[Hình I.1 – Kiến trúc khái quát của mạng nơ-ron sâu [1] 6](#_Toc121394112)

[Hình II.1 – Kiến trúc của CNN model (VGG-16) 8](#_Toc121394113)

[Hình II.2 – Cấu trúc của bản ghi mẫu 16x16 9](#_Toc121394114)

[Hình II.3 – Công thức Normalization tính giá trị cho mỗi đặc trưng 10](#_Toc121394115)

[Hình II.4 – Ví dụ bản ghi đã được chuyển đổi thành định dạng ảnh 10](#_Toc121394116)

[Hình III.1 – Quy trình huấn luyện và thử nghiệm mô hình với SMOTE-ENN [4] 12](#_Toc121394117)

**MỤC LỤC BẢNG**

[Bảng II.1 – Số lượng bản ghi của tập dữ liệu huấn luyện và thử 9](#_Toc121394107)

**THUẬT NGỮ**

**Activation Function** – Hàm kích hoạt

**Artificial Intelligence** – Trí tuệ nhân tạo

**Bias** – Độ lệch giá trị trung bình

**Convolutional Neural Network** – Mạng nơ-ron tích chập

**Deep Convolutional Neural Network** – Mạng nơ-ron tích chập sâu

**Intrusion Detection System** – Hệ thống phát hiện xâm nhập

**Grayscale** – Thước xám

**Neural Network** – Mạng nơ-ron

**Overfitting** – Quá khớp

**Raw Data** – Dữ liệu thô

**Render** – Kết xuất

**Sampling** – Lấy mẫu

**Stride** – Bước nhảy

**Weight** – Trọng số

**Federated Learning / Collaborative Learning** – Học cộng tác

# Nghiên cứu cơ sở lý thuyết

## Mạng nơ-ron – Neural Network

Là phương thức trong trí tuệ nhân tạo dùng để huấn luyện máy tính xử lý và tính toán dữ liệu thô. Sử dụng các nơ-ron hoặc node tạo thành cấu trúc phân lớp có tính thích ứng cao từ đó liên tục cải thiện để cho ra kết quả tốt nhất. Được sử dụng để giải quyết những vấn đề phức tạp như dự đoán, nhận diện, … [1]

Một kiến trúc mạng nơ-ron sẽ bao gồm ba loại layer : [2]

* Input Layer : Là dữ liệu đầu vào được cho vào model. Số lượng phần tử nơ-ron ở layer này đại diện cho số lượng feature của dữ liệu.
* Hidden Layer(s) : Dữ liệu đầu vào sẽ được đưa từ Input Layer sang các Hidden Layer. Từng layer có thể có số lượng nơ-ron khác nhau. Đây là nơi output từ layer trước sẽ được nhân với trọng số (weight) của layer đó, tiếp đó là độ lệch giá trị trung bình (bias) và hàm kích hoạt (Activation Function).
* Output Layer : Là nơi dữ liệu đã được xử lý đi ra. Tương tự với Input Layer, chỉ có duy nhất một output layer duy nhất với nhiều phần tử nơ-ron.



Hình . – Kiến trúc khái quát của mạng nơ-ron sâu [1]

## Mạng nơ-ron tích chập – Convolutional Neural Network

Là một dạng của mạng nơ-ron bao gồm nhiều lớp nơ-ron mà trong đó có ít nhất một lớp tích chập mang cho mình nhiệm vụ xử lý dữ liệu được đưa vào. Một mạng nơ-ron tích chập sẽ bao gồm những layer như sau : [3]

* **Input Layer** **:** Như Input Layer của mạng nơ-ron, chỉ có một Input Layer duy nhất là nơi dữ liệu thô được đưa vào.
* **Convolution Layer(s) :** Bao gồm các bộ lọc. Ở mỗi nơ-ron tích chập, dữ liệu đầu ra có chiều dài và độ cao nhỏ hơn so với dữ liệu thô nhưng cùng độ sâu (Ví dụ như dữ liệu thô là ảnh định dạng RGB thì độ sâu sẽ là 3 cho dữ liệu đầu ra). Vì thế nên khi dữ liệu thô đi qua mỗi lớp (layer) tích chập ta sẽ có dữ liệu đầu ra có cùng chiều dài và độ cao, độ sâu của dữ liệu đầu ra sẽ được quyết định bởi số bộ lọc và bước nhảy ở lớp tích chập.

Nếu như một mạng nơ-ron tích chập có nhiều lớp tích chập thì nó sẽ trở thành mạng nơ-ron tích chập sâu – Deep Convolutional Neural Network.

* **Activation Function Layer :** Dữ liệu từ lớp tính chập đi qua đây sẽ được áp dụng các hàm tính phi tuyến tính như ReLU, Sigmoid, Tanh, …
* **Pool Layer :** Được đưa vào mạng nơ-ron tính chập nhằm giảm kích thước của khối nhằm tăng khả năng tính toán cho những lớp sau và đồng thời tránh vấn đề quá khớp. Thường có hai loại là max-pooling và average-pooling (hay còn gọi là mean pooling).
* **Fully-connected Layer :** Kế thừa dữ liệu từ các lớp trước, mỗi nơ-ron sẽ được kết nối với toàn bộ các nơ-ron trước. Thường nằm ở cuối mạng tính toán các điểm số (score) của lớp như độ chính xác của từng lớp.

## Tâp dữ liệu không đều – Imbalanced Dataset

Sự mất cân bằng trong tập dữ liệu xảy ra trong quá trình phân lớp, khi mà chệnh lệch giữa số bản ghi của lớp thiểu số và lớp đa số quá lớn. Dẫn đến khi thuật toán được huấn luyện với những bộ dữ liệu mất cân bằng, kết quả sẽ có khuynh hướng thiên về lớp đa số hơn lớp thiểu số và làm ảnh hưởng đến hiệu quả chung.[12]

Và qua những số liệu về tập dữ liệu UNSW-NB15, ta có thể thấy rằng có sự chênh lệch rất lớn về số bản ghi của lớp hoạt động mạng bình thường (Normal) và lớp hành vi tấn công sâu máy tính (Worms) (Tỉ lệ lên đến 1000:1). [7][8]

Điều này dẫn đến sự mất cân bằng giữa lớp đa số và thiểu số trong tập dữ liệu.

Từ đó nhóm tìm hiểu thêm về những cách tiếp cận mới nhằm cải thiện hiệu quả của mô hình DCNN ở phần trước. Một trong số các cách tiếp cận nhằm cải thiện nhược điểm của tập dữ liệu này là việc chúng ta tiến hành Resampling. [4][12]

## SMOTE – Synthetic Minority Over-sampling

(Sẽ bổ sung)

## ENN – Edited Nearest Neighbors

(Sẽ bổ sung)

## Máy học cộng tác – Federated Learning

(Sẽ bổ sung)

# Khảo sát các nghiên cứu liên quan

## Model cho mạng nơ-ron tích chập

Dựa trên một số nghiên cứu về hệ thống phát hiện xâm nhập trên nền tảng học sâu, nhóm sẽ sử dụng VGG-16 – Một model mạng nơ-ron tích chập sâu.

VGG-16 được Karen Simonyan và Andrew Zisserman giới thiệu vào 2014 ở cuộc thi ILSVRC 2014 [5] qua bài báo [6].

Về kiến trúc của VGG-16, Input dành cho model này sẽ là một ảnh RGB size 224x224. Sẽ có tất cả 16 lớp, 5 lớp tích chập được theo sau bởi max pooling. Tất cả các lớp tích chập sẽ sử dụng bộ lọc có kích thước 3x3 (hoặc 1x1). Thứ tự các lớp được thể hiện qua hình dưới đây.



Hình . – Kiến trúc của CNN model (VGG-16)

## Tập dữ liệu

Tập dữ liệu UNSW-NB15 được tạo bởi UNSW Canberra nhằm mô tả hoạt động mạng hỗn hợp bao gồm hoạt động bình thường và các hành vi tấn công mạng. Tập dữ liệu bao gồm 9 phân loại tấn công và 49 đặc trưng. Số lượng bản ghi của mỗi tập dữ liệu con và tập dữ liệu mẹ được thể hiện qua bảng sau.

Đối với hai tập dữ liệu con là tập dữ liệu con dành cho huấn luyện và tập dữ liệu con dành cho thử. Các bản ghi về hoạt động mạng bình thường được trích ra một phần trong hơn 2 triệu bản ghi. Về phần các bản ghi lại hành vi tấn công mạng được chia ra cho cả hai tập dữ liệu con với tỉ lệ 7:3. [7][8]

|  |  |  |
| --- | --- | --- |
| Tên lớp | Tập dữ liệu con dành cho huấn luyện | Tập dữ liệu con dành cho thử |
| Normal | 56,000 | 37,000 |
| Generic | 40,000 | 18,871 |
| Exploits | 33,393 | 11,132 |
| Fuzzers | 18,184 | 6,062 |
| DoS | 12,264 | 4,089 |
| Reconnaissance | 10,491 | 3,496 |
| Analysis | 2,000 | 677 |
| Backdoor | 1,746 | 583 |
| Shellcode | 1,133 | 378 |
| Worms | 130 | 44 |
| Tổng cộng | 175,341 | 82,332 |

Bảng . – Số lượng bản ghi của tập dữ liệu huấn luyện và thử

## Tiền xử lý dữ liệu

Sử dụng mã hóa One-Hot và kỹ thuật Normalization, ta có thể biến đổi các bản ghi của tập dữ liệu sang loại dữ liệu đầu vào phù hợp với mạng nơ-ron tích chập.

Đầu tiên tập dữ liệu gốc sẽ được tách từng hàng ngang (mỗi hàng ngang tương đương với một bản ghi). Mỗi bảng ghi đã được tách ra sẽ chuyển về dạng bảng với mỗi ô là một đặc tính, từ đó tạo nên bảng vuông có độ dài là 16 đặc tính tương đương với 16 ô. Những ô không chứa dữ liệu của bất kỳ đặc tính nào sẽ được pad để đảm bảo kích thích của ảnh sau này. [9]

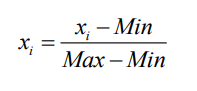
Table

Description automatically generated

Hình . – Cấu trúc của bản ghi mẫu 16x16

Mã hóa one-hot sẽ đưa giá trị của các đặc trưng từ giá trị dạng số trở thành giá trị thước xám. Với các ô không chứa dữ liệu (pad) sẽ được biến đổi thành các pixel màu đen (có giá trị bằng 255).

Sau cùng bằng kỹ thuật Normalization (Hình II.3) thì tất cả các đặc trưng (dịch vụ, giao thức, trạng thái, …) của bản ghi sẽ được biến đổi thành 1 nếu như có xuất hiện và 0 nếu như đặc trưng nào không xuất hiện trong bản ghi. Ở hình dưới đây, xi chính là giá trị của đặc trưng, Max và Min thể hiện giá trị cao nhất và thấp nhất của đặc trưng x.



Hình . – Công thức Normalization tính giá trị cho mỗi đặc trưng

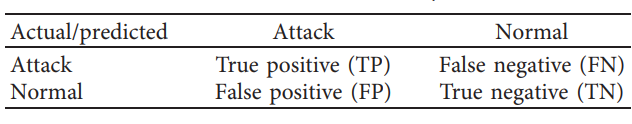
Sau cùng từng bản ghi 16x16 sẽ được chuyển đổi thành hình ảnh có kích thước tương tự là 16x16. Như ở hình dưới đây lần lượt là bản ghi của lớp analysis, lớp backdoor và lớp dos của tập dữ liệu dùng cho việc thử.



Hình . – Ví dụ bản ghi đã được chuyển đổi thành định dạng ảnh

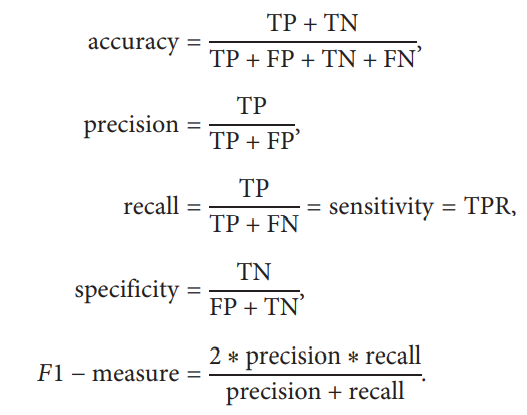
## Phương pháp đánh giá hiệu quả

Vì các đặc trưng mang giá trị nhị phân, confusion matrix được sử dụng để đánh giá độ hiệu quả của model. [7]



Bảng . – Confusion Matrix cho việc phân loại

Và từ metric của Confusion Matrix, ta có những công thức như sau.



Trong đó :

* TP là bản ghi cuộc tấn công được phân loại đúng
* FP là bản ghi hoạt động bình thường được phân loại thành tấn công
* TN là bản ghi hoạt động bình thường được phân loại đúng
* FN là bản ghi cuộc tấn công được phân loại thành hoạt động bình thường

# Thực nghiệm

## Trên mạng nơ-ron tích chập sâu

Về tập dữ liệu thực nghiệm, nhóm sử dụng tập dữ liệu UNSW-NB15 đã qua quá trình tiền xử lý nhằm cho ra dữ liệu đầu vào phù hợp với DCNN.

### Local model

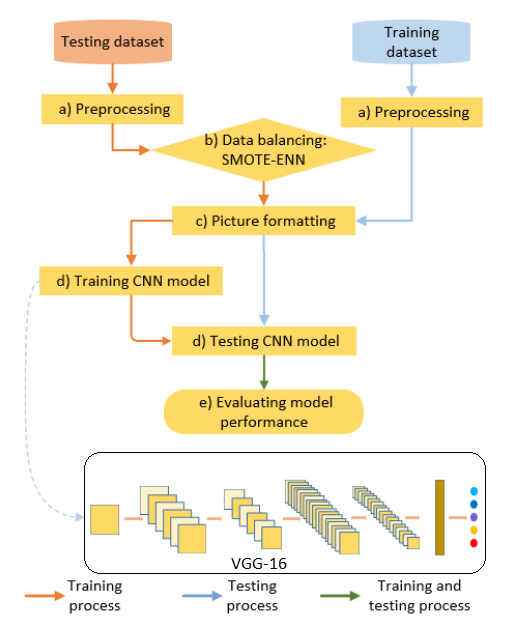
Ở Local model, model VGG-16 sẽ được huấn luyện ngay chính trên thiết bị sinh dữ liệu. Mô hình mạng nơ-ron tích chập này có cấu tạo như sau

(Đánh giá chi tiết sẽ được bổ sung)

## Cải thiện với SMOTE-ENN

### Quy trình cải thiện tập dữ liệu

Nhằm khắc phục hạn chế về sự cân bằng của tập dữ liệu UNSW-NB15. Với phương pháp lấy mẫu hỗn hợp (Bao gồm cả kỹ thuật Over-sampling và Under-sampling). Quy trình xử lý và cải thiện tập dữ liệu có bổ sung như sau. [4]



Hình . – Quy trình huấn luyện và thử nghiệm mô hình với SMOTE-ENN [4]

Tương tự với quy trình cũ, mã hóa One-Hot và kỹ thuật Normalization sẽ được sử dụng cho tiền xử lý dữ liệu nhằm chuyển đổi tập dữ liệu csv sang dạng hình ảnh cho mạng nơ-ron tích chập.

Sau đó tập dữ liệu dùng cho việc huấn luyện mô hình CNN sẽ được lấy mẫu lại bằng kỹ thuật Over-sampling kết hợp với Under-sampling (SMOTE + ENN). Sau khi từng bản ghi của tập dữ liệu này sẽ được chuyển đổi thành hình ảnh có kích thước 16x16, tập dữ liệu sẽ sẵn sàng cho việc huấn luyện mô hình.

Quy trình chuyển đổi tập dữ liệu sang định dạng hình ảnh vẫn được giữ nguyên với tập dữ liệu dùng cho việc kiểm thử. Sau khi tiền xử lý các bản ghi với mã hóa One-Hot và kỹ thuật Normalization, từng bản ghi sẽ được chuyển đổi sang định dạng hình ảnh mà không cần phải qua kỹ thuật lấy mẫu hỗn hợp SMOTE + ENN. Sau đó thì tập dữ liệu con này đã sẵn sàng cho việc kiểm thử mô hình đã được huấn luyện.

**References**

[1] : ibm.com/cloud/learn/neural-networks

[2] : aws.amazon.com/what-is/neural-network

[3] : Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, *“A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,”* IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 12. Institute of Electrical and Electronics Engineers (IEEE), pp. 6999–7019, Dec. 2022. doi: 10.1109/tnnls.2021.3084827.

[4] X. Zhang, J. Ran, and J. Mi, *“An Intrusion Detection System Based on Convolutional Neural Network for Imbalanced Network Traffic,”* 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT). IEEE, Oct. 2019. doi: 10.1109/iccsnt47585.2019.8962490.

[5] image-net.org/challenges/LSVRC/2014/

[6] K. Simonyan and A. Zisserman, *“Very Deep Convolutional Networks for Large-Scale Image Recognition.”* arXiv, 2014. doi: 10.48550/ARXIV.1409.1556.

[7] N. Moustafa and J. Slay, *“The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set,”* Information Security Journal: A Global Perspective, vol. 25, no. 1–3. Informa UK Limited, pp. 18–31, Jan. 11, 2016. doi: 10.1080/19393555.2015.1125974.

[8] N. Moustafa and J. Slay, *“UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),”* 2015 Military Communications and Information Systems Conference (MilCIS). IEEE, Nov. 2015. doi: 10.1109/milcis.2015.7348942.

[9] D. A. Noever and S. E. M. Noever, *“Image Classifiers for Network Intrusions.”* arXiv, 2021. doi: 10.48550/ARXIV.2103.07765.

[10] D. Ramyachitra and P. Manikandan. *“Imbalanced Dataset Classification And Solutions : A Review.”* 2014.

[11] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, *“An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset,”* Computer Networks, vol. 177. Elsevier BV, p. 107315, Aug. 2020. doi: 10.1016/j.comnet.2020.107315.

[12] M. Koziarski, *“Two-Stage Resampling for Convolutional Neural Network Training in the Imbalanced Colorectal Cancer Image Classification,”* 2021 International Joint Conference on Neural Networks (IJCNN). IEEE, Jul. 18, 2021. doi: 10.1109/ijcnn52387.2021.9533998.