**ỨNG DỤNG CỦA DEEP LEARNING TRONG PHÁT HIỆN TẤN CÔNG MẠNG TRÊN TẬP DỮ LIỆU CICIOT2023**

**Nguyễn Tuấn Trọng**

Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Kỹ thuật Vĩnh Long

[21004235@st.vlute.edu.vn](mailto:21004235@st.vlute.edu.vn)

***Tóm tắt*** *– Sự phát triển của Internet vạn vật (IoT) đã làm gia tăng các mối đe dọa an ninh mạng, đặc biệt là các cuộc tấn công nhắm vào thiết bị IoT như đánh cắp dữ liệu, DDoS và cấy mã độc. Bài báo này đề xuất một phương pháp sử dụng các mô hình học sâu, bao gồm DCN-V2, FT-Transfosmer, TabNet, TabR và Node-Gam để phát hiện tấn công mạng trên tập dữ liệu CICIoT2023. Các mô hình được huấn luyện trên nhiều kịch bản 2 nhãn, 8 nhãn và 34 nhãn, cuối cùng đánh giá nhằm tối ưu hóa độ chính xác, giảm thiểu cảnh báo sai và nâng cao khả năng ứng dụng thực tế. Kết quả thực nghiệm cho thấy phương pháp đề xuất đạt độ chính xác cao, vượt trội so với các phương pháp truyền thống. Nghiên cứu cũng đề xuất các cải tiến để tăng cường hiệu quả bảo mật cho hệ thống IOT, hỗ trợ phát hiện kịp thời các mối đe dọa mạng.*

***Từ khóa*** *– Học sâu, IoT, CICIoT2023, phát hiện tấn công mạng, DCN-V2, FT-Transfosmer, TabNet, TabR, Node-Gam.*

**I. GIỚI THIỆU**

***A. Giới thiệu bài toán***

Internet vạn vật (IoT) đã trở thành nền tảng quan trọng trong cách mạng công nghiệp 4.0, kết nối hàng tỷ thiết bị thông minh trong các lĩnh vực như công nghiệp, y tế, giao thông và nhà thông minh. Theo Statista, số lượng thiết bị IoT toàn cầu dự kiến đạt 30,9 tỷ vào năm 2025 [1]. Các thiết bị này liên tục thu thập, xử lý và truyền dữ liệu qua Internet, mang lại hiệu quả quản lý và tối ưu hóa quy trình. Tuy nhiên, sự mở rộng nhanh chóng của IoT cũng kéo theo các thách thức lớn về an ninh mạng.

Do hoạt động liên tục trên Internet, các thiết bị IoT trở thành mục tiêu hấp dẫn cho các cuộc tấn công mạng. Kaspersky ghi nhận hơn 1,5 tỷ cuộc tấn công nhắm vào thiết bị IoT trong nửa đầu năm 2021 [2]. Các hình thức tấn công phổ biến bao gồm đánh cắp dữ liệu, kiểm soát trái phép, tấn công từ chối dịch vụ phân tán (DDoS) và cấy mã độc. Vụ tấn công Mirai Botnet năm 2016 là một ví dụ điển hình, khi hơn 100.000 thiết bị IoT bị xâm nhập để tạo mạng botnet, gây ra các cuộc tấn công DDoS với lưu lượng 1,2 Tbps, làm gián đoạn nhiều dịch vụ trực tuyến [3]. Những sự cố này nhấn mạnh nhu cầu cấp thiết về các giải pháp bảo mật tiên tiến để bảo vệ hệ thống IoT trước các mối đe dọa ngày càng tinh vi.

Các phương pháp bảo mật truyền thống, như phát hiện dựa trên chữ ký hoặc quy tắc, thường không hiệu quả trước các cuộc tấn công mới và biến đổi. Theo IBM Security năm 2024, các tổ chức sử dụng phương pháp truyền thống mất trung bình 204 ngày để phát hiện vi phạm dữ liệu, và các mối đe dọa không khớp với chữ ký đã biết dễ dàng vượt qua hệ thống bảo mật [4]. Điều này đòi hỏi các giải pháp bảo mật thông minh hơn, có khả năng học hỏi và thích ứng với các mẫu tấn công mới.

Học sâu (Deep Learning) đã chứng minh tiềm năng vượt trội trong an ninh mạng, đặc biệt trong phát hiện tấn công. Các mô hình như Mạng nơ-ron tích chập (CNN), Mạng nơ-ron hồi quy (RNN) và Transformer có thể trích xuất đặc trưng phức tạp từ dữ liệu lớn và phát hiện các mẫu bất thường với độ chính xác cao [5]. Nghiên cứu từ IEEE năm 2021 cho thấy học sâu đạt độ chính xác 98% trong phát hiện xâm nhập mạng IoT [6]. Tập dữ liệu CICIoT2023, phát triển bởi Trung tâm Nghiên cứu An ninh mạng Canada, cung cấp dữ liệu thực tế về 33 loại tấn công mạng nhắm vào thiết bị IoT, là nguồn tài nguyên lý tưởng để huấn luyện và đánh giá mô hình học sâu [7].

Bài báo này đề xuất một phương pháp sử dụng học sâu để phát hiện tấn công mạng trên tập dữ liệu CICIoT2023, nhằm xây dựng hệ thống bảo mật tự động, chính xác và hiệu quả. Phương pháp này tập trung vào tối ưu hóa các mô hình học sâu để nâng cao khả năng phát hiện, giảm thiểu rủi ro và bảo vệ hệ thống IoT trước các mối đe dọa mạng.

***B. Những nghiên cứu liên quan***

Trong những năm gần đây, nhiều nghiên cứu quốc tế đã ứng dụng học sâu để phát hiện tấn công mạng trong môi trường IoT, với tập dữ liệu CICIoT2023 là nền tảng thực tế để đánh giá. Ferrag và cộng sự [8] đề xuất mô hình kết hợp CNN và RNN trên BoT-IoT và CICIoT2023, đạt độ chính xác 95,8% trong phát hiện tấn công DDoS, nhưng hiệu suất giảm với các cuộc tấn công zero-day do dữ liệu huấn luyện hạn chế. Ullah và cộng sự [9] sử dụng Transformer với học chuyển giao trên CICIoT2023, đạt độ chính xác 96,3% và AUC 0,972, hiệu quả với tấn công MITM, dù chi phí tính toán cao. Al-Qarafi và cộng sự [10] áp dụng Faster R-CNN trên CICIoT2023, đạt độ chính xác 94,7% trong nhận diện brute-force và scanning, nhưng yêu cầu tiền xử lý phức tạp. Neto và cộng sự [11] kết hợp CNN để phân loại tấn công DDoS và reconnaissance trên CICIoT2023, đạt độ chính xác 95,6%, tuy nhiên hiệu suất giảm khi gặp dữ liệu nhiễu. Ullah và cộng sự [12] sử dụng ResNet-18 với tăng cường dữ liệu trên CICIoT2023, đạt độ chính xác 96,9% và F1-score 0,958, nhưng chi phí tính toán hạn chế triển khai trên thiết bị IoT yếu. Al-Haija và cộng sự [13] phát triển IoT-IDCS-CNN trên NSL-KDD, đạt độ chính xác 98,2% trong phân loại đa lớp, có tiềm năng áp dụng cho CICIoT2023, dù yêu cầu tài nguyên lớn. Popoola và cộng sự [14] kết hợp Autoencoder và LSTM trên BoT-IoT, đạt độ chính xác 97,2%, hiệu quả với brute-force, nhưng cần cân bằng dữ liệu để xử lý dữ liệu mất cân bằng. Sun và cộng sự [15] sử dụng CNN kết hợp

LSTM trên UNSW-NB15, đạt độ chính xác 98,1%, có thể mở rộng sang CICIoT2023, nhưng cần dữ liệu huấn luyện lớn. Haq và Khan [16] áp dụng DNN trên BoT-IoT, đạt độ chính xác 97,5% với kỹ thuật SMOTE, nhưng khó phân biệt tấn công có đặc trưng gần lưu lượng bình thường. So với các nghiên cứu trên, bài báo này tập trung vào việc tối ưu hóa các mô hình học sâu tiên tiến, bao gồm DCN-V2, FT-Transformer, TabNet, TabR và Node-GAM, để phát hiện tấn công mạng trên tập dữ liệu CICIoT2023. Nghiên cứu triển khai các nhiệm vụ phân loại đa dạng, bao gồm phân loại nhị phân (2 nhãn), phân loại 8 nhãn và phân loại chi tiết 34 nhãn.

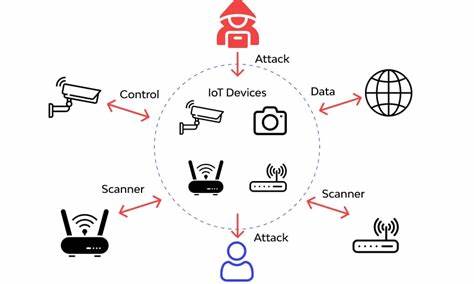
**II. MỘT SỐ VẤN ĐỀ LIÊN QUAN**

**A. Tân công mạng**

Tấn công mạng là các hành vi cố ý nhằm xâm nhập, làm gián đoạn hoặc khai thác trái phép hệ thống máy tính, mạng viễn thông, hoặc thiết bị kết nối, với mục tiêu đánh cắp dữ liệu, phá hoại hoạt động, hoặc gây thiệt hại cho người dùng và tổ chức. Theo Sophos 2024 State of Ransomware, các cuộc tấn công mạng, đặc biệt là ransomware, tăng 37% trong năm 2024, với thiết bị IoT chiếm 15% điểm xâm nhập [17]. Các thiết bị IoT, như camera an ninh, bộ định tuyến, và cảm biến thông minh, thường thiếu biện pháp bảo mật cơ bản, trở thành mục tiêu lý tưởng cho các cuộc tấn công.

Trong bối cảnh IoT, tấn công mạng diễn ra dưới nhiều hình thức với mức độ nghiêm trọng khác nhau. Báo cáo Check Point 2024 Cyber Security Report chỉ ra rằng các quốc gia phát triển về hạ tầng số như Hoa Kỳ và Tây Âu chiếm 65% các cuộc tấn công mạng toàn cầu, trong khi các khu vực như Nam Á có tỷ lệ thấp hơn do hạn chế về kết nối [18]. Các yếu tố dẫn đến tấn công mạng bao gồm lỗ hổng bảo mật phần mềm, với 78% vụ tấn công năm 2024 khai thác lỗ hổng tồn tại hơn một năm [19], thiếu nhận thức an ninh mạng của người dùng [20], và sự gia tăng của thiết bị IoT không được bảo vệ đầy đủ [21]. Tập dữ liệu CICIoT2023, mô phỏng 33 loại tấn công IoT, là nguồn dữ liệu thực tế để nghiên cứu và phát triển các giải pháp phát hiện dựa trên học sâu [22].

Các loại tấn công mạng phổ biến trong IoT bao gồm:

* Tấn công từ chối dịch vụ phân tán (DDoS)
* Tấn công từ chối dịch vụ (DoS)
* Tấn công Man-in-the-Middle (MITM)
* Tấn công brute-force
* Tấn công thăm dò (Reconnaissance)
* Tấn công khai thác lỗ hổng (Exploitation)
* Tấn công giả mạo (Spoofing)

**Hình 1 :** Minh họa các cuộc tấn công mạng nhắm vào thiết bị IoT

**B. Các mạng học sâu được dùng trong đề tài**

Nghiên cứu này sử dụng các mô hình tiên tiến gồm DCN-V2, FT-Transformer, TabNet, TabR, AutoInt và Node-GAM. Các mô hình này được lựa chọn nhờ khả năng xử lý hiệu quả dữ liệu dạng bảng, học các tương tác đặc trưng phức tạp, và phù hợp với các nhiệm vụ phân loại đa nhãn (2 nhãn, 8 nhãn, 34 nhãn) trong bối cảnh IoT.

1. **Mạng FT-Transformer**

FT-Transformer (Feature Tokenizer + Transformer) [36] là một kiến trúc học sâu được thiết kế để xử lý dữ liệu dạng bảng, phù hợp cho bài toán phân loại tấn công mạng. Mô hình chuyển đổi các đặc trưng đầu vào thành token và áp dụng các tầng Transformer để học tương tác đặc trưng. FT-Transformer bao gồm hai thành phần chính: Feature Tokenizer (mã hóa đặc trưng thành vector nhúng) và khối Transformer (sử dụng tự chú ý đa đầu để học mối quan hệ phi tuyến). Token [CLS] được thêm vào để tổng hợp thông tin toàn cục, và biểu diễn cuối cùng của token này được dùng để dự đoán. Nghiên cứu này sử dụng FT-Transformer để phân loại các loại tấn công như DDoS, MITM và brute-force, tận dụng khả năng học tương tác đặc trưng phức tạp.

FT-Transformer có nhiều ưu điểm: tự động học các mối quan hệ phi tuyến giữa các đặc trưng, hỗ trợ đồng thời

đặc trưng số và phân loại mà không cần tiền xử lý phức tạp, và khả năng mở rộng nhờ huấn luyện song song trên GPU. Điều này giúp mô hình đạt hiệu suất cao trên tập dữ liệu lớn như CICIoT2023.

**2. Mạng DCN-V2**

DCN-V2 (Deep & Cross Network V2) [40] là phiên bản cải tiến của DCN, được thiết kế để xử lý dữ liệu bảng trong các tác vụ phân loại. Mô hình kết hợp Cross Network (học tương tác đặc trưng rõ ràng) và Deep Network (học tương tác ngầm) để nâng cao hiệu suất. DCN-V2 sử dụng ma trận trọng số trong Cross Network và kỹ thuật phân rã hạng thấp để giảm chi phí tính toán. Trong nghiên cứu này, DCN-V2 được áp dụng để phân loại các loại tấn công mạng, tận dụng khả năng học các tương tác đặc trưng cấp cao.

DCN-V2 có ưu điểm vượt trội: học hiệu quả cả tương tác rõ ràng và ngầm, giảm chi phí tính toán nhờ phân rã hạng thấp, và dễ tích hợp vào các hệ thống thực tiễn. Mô hình này phù hợp với tập dữ liệu lớn như CICIoT2023, nơi cần xử lý hàng triệu mẫu lưu lượng mạng.

**3. Mạng TabNet**

TabNet [38] là một kiến trúc học sâu dành cho dữ liệu bảng, sử dụng cơ chế chú ý tuần tự để chọn lọc đặc trưng tại mỗi bước ra quyết định. TabNet bao gồm các bước xử lý tuần tự, mỗi bước tạo biểu diễn đặc trưng và đóng góp vào dự đoán cuối cùng. Mô hình hỗ trợ học tự giám sát trên dữ liệu không nhãn, cải thiện hiệu suất khi dữ liệu có nhãn khan hiếm. Trong nghiên cứu này, TabNet được sử dụng để phân loại các cuộc tấn công mạng, tập trung vào các đặc trưng quan trọng như lưu lượng gói tin và giao thức.

TabNet nổi bật với khả năng cạnh tranh với các mô hình cây quyết định như XGBoost, đồng thời cung cấp tính giải thích nhờ cơ chế chú ý. Mô hình giảm hiện tượng quá khớp và đạt hiệu suất cao trên dữ liệu bảng như CICIoT2023.

**4. Mạng Node-GAM**

Node-GAM (Neural Generalized Additive Model) [41] là một mô hình học sâu kết hợp tính giải thích của Generalized Additive Models (GAMs) với hiệu suất của mạng nơ-ron sâu. Mô hình sử dụng cây quyết định vô thức phân biệt và cơ chế chú ý để học các hàm phi tuyến độc lập cho từng đặc trưng. Node-GAM được áp dụng trong nghiên cứu này để phân loại tấn công mạng, đảm bảo cả độ chính xác và khả năng giải thích kết quả.

Node-GAM có ưu điểm: cung cấp tính giải thích cao, vượt trội trên dữ liệu lớn so với GAM truyền thống, và hỗ trợ học tự giám sát để cải thiện hiệu suất trên dữ liệu khan hiếm. Mô hình phù hợp để phân tích các đặc trưng lưu lượng mạng trong CICIoT2023.

**5. Mạng TabR**

TabR (Tabular Deep Learning Meets Nearest Neighbors) [42] là một mô hình học sâu tích hợp mạng nơ-ron với cơ chế k-Nearest Neighbors (k-NN). TabR truy xuất các mẫu tương tự từ dữ liệu huấn luyện để tăng cường dự đoán, kết hợp nhúng đặc trưng và thông tin láng giềng qua mạng truyền thẳng. Trong nghiên cứu này, TabR được sử dụng để phân loại các loại tấn công mạng, tận dụng khả năng khai thác mẫu tương tự trong dữ liệu.

TabR có ưu điểm: đạt hiệu suất cao nhờ cơ chế truy xuất láng giềng, đơn giản và nhanh, đồng thiết kế mô-đun dễ tích hợp vào các hệ thống thực tiễn. Mô hình này phù hợp với các tập dữ liệu lớn như CICIoT2023.

**C. Đánh giá mô hình bài toán**

Trong bài toán phát hiện tấn công mạng trên tập dữ liệu CICIoT2023, việc lựa chọn các độ đo đánh giá phù hợp là rất quan trọng để so sánh hiệu suất của các mô hình học sâu (DCN-V2, FT-Transformer, TabNet, TabR, AutoInt, Node-GAM) trong các nhiệm vụ phân loại đa nhãn (2 nhãn, 8 nhãn, 34 nhãn). Các độ đo được sử dụng gồm Accuracy, Precision, Recall, F1-Score và Loss, giúp đánh giá toàn diện từ độ chính xác tổng thể đến khả năng phát hiện từng lớp tấn công.

**1. Độ chính xác (Accuracy)**

Accuracy là tỷ lệ phần trăm các mẫu được mô hình dự đoán đúng so với tổng số mẫu, cung cấp cái nhìn tổng quan về hiệu suất phân loại. Công thức tính Accuracy được biểu diễn như sau:

Trong đó :

* **True Positive (TP)**: Số mẫu thuộc lớp dương được dự đoán đúng là lớp dương.
* **True Negative (TN)**: Số mẫu thuộc lớp âm được dự đoán đúng là lớp âm.
* **False Positive (FP)**: Số mẫu thuộc lớp âm nhưng bị dự đoán sai thành lớp dương.
* **False Negative (FN)**: Số mẫu thuộc lớp dương nhưng bị dự đoán sai thành lớp âm.

**2. Độ chính xác từng lớp (Precision)**

Precision đo lường tỷ lệ các mẫu được dự đoán là tấn công thực sự đúng, phản ánh độ tin cậy của dự đoán dương tính. Công thức tính Precision được biểu diễn như sau:

Precision đặc biệt quan trọng trong bài toán phát hiện tấn công mạng, nơi cần giảm thiểu các cảnh báo sai (FP) để tránh làm gián đoạn hệ thống không cần thiết.

**3. Độ phủ (Recall)**

Recall đo lường tỷ lệ các mẫu tấn công được phát hiện đúng, phản ánh khả năng bao quát của mô hình trên lớp dương. Công thức tính Recall được biểu diễn như sau:

Recall rất quan trọng trong bối cảnh IoT, vì bỏ sót một cuộc tấn công (FN) có thể gây hậu quả nghiêm trọng cho hệ thống.

**4. F1-Score**

F1-Score là trung bình điều hòa của Precision và Recall, cân bằng giữa độ chính xác và độ phủ, đặc biệt hữu ích khi dữ liệu mất cân bằng. Công thức tính F1-Score được biểu diễn như sau:

F1-Score giúp đánh giá hiệu suất trên từng lớp tấn công (DDoS, MITM, brute-force, v.v.), đảm bảo mô hình không thiên vị giữa Precision và Recall.

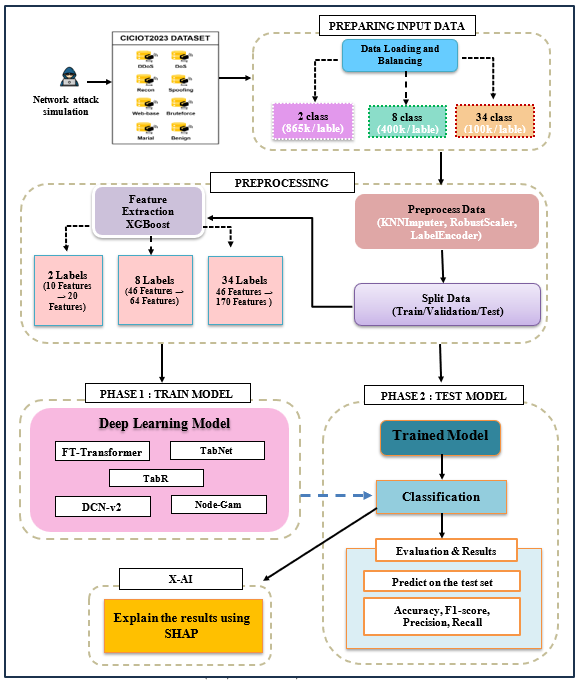
**5. Độ mất mát (Loss)**

Loss đo lường mức độ sai lệch giữa dự đoán của mô hình và nhãn thực tế, được sử dụng để tối ưu hóa mô hình trong quá trình huấn luyện. Trong bài toán phân loại đa nhãn, hàm mất mát cross-entropy được sử dụng. Công thức tính Loss được biểu diễn như sau:

Trong đó :

* N: Số mẫu trong tập dữ liệu.
* C : Số lớp.
* pi(c): Xác suất thực tế của mẫu i thuộc lớp c (thường là 1 nếu đúng lớp, 0 nếu không).
* qi(c): Xác suất dự đoán của mô hình cho mẫu i thuộc lớp c.

**III. PHƯƠNG PHÁP ĐỀ XUẤT**

**Hình 2 :** Phương pháp đề xuất

Phương pháp đề xuất được thiết kế để phát hiện các cuộc tấn công mạng trên tập dữ liệu CICIoT2023, sử dụng các kỹ thuật học sâu (deep learning) nhằm phân loại các mẫu dữ liệu theo ba mức độ chi tiết: 2 nhãn (phân biệt dữ liệu bình thường và tấn công), 8 nhãn (phân loại 8 loại tấn công khác nhau), và 34 nhãn (phân loại chi tiết 34 loại tấn công). Quy trình được chia thành hai giai đoạn chính: giai đoạn huấn luyện (Phase 1: Train Model) và giai đoạn kiểm thử (Phase 2: Test Model). Đầu tiên, dữ liệu được tải và cân bằng (data loading and balancing) để đảm bảo tỷ lệ hợp lý giữa các nhãn.

**IV. KẾT QUẢ THỰC NGHIỆM**

**A. Môi trường cài đặt và tập dữ liệu**

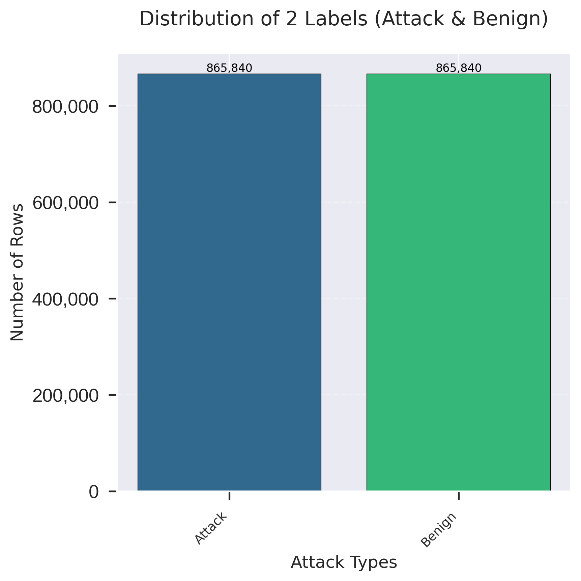
**1. Môi trường cài đặt**

Đề tài được triển khai bằng ngôn ngữ Python và chạy trên nền tảng Kaggle với gói tài nguyên Kaggle Notebooks, sử dụng bộ vi xử lý CPU 4 nhân (Intel® Xeon® 2.30 GHz), bộ nhớ RAM 16GB, và bộ xử lý đồ họa GPU NVIDIA Tesla T4 với dung lượng RAM 15GB. Các thư viện hỗ trợ huấn luyện mô hình học sâu bao gồm PyTorch v2.0, scikit-learn v1.3, và pandas v2.1, cùng với các công cụ hỗ trợ tiền xử lý dữ liệu và đánh giá mô hình.

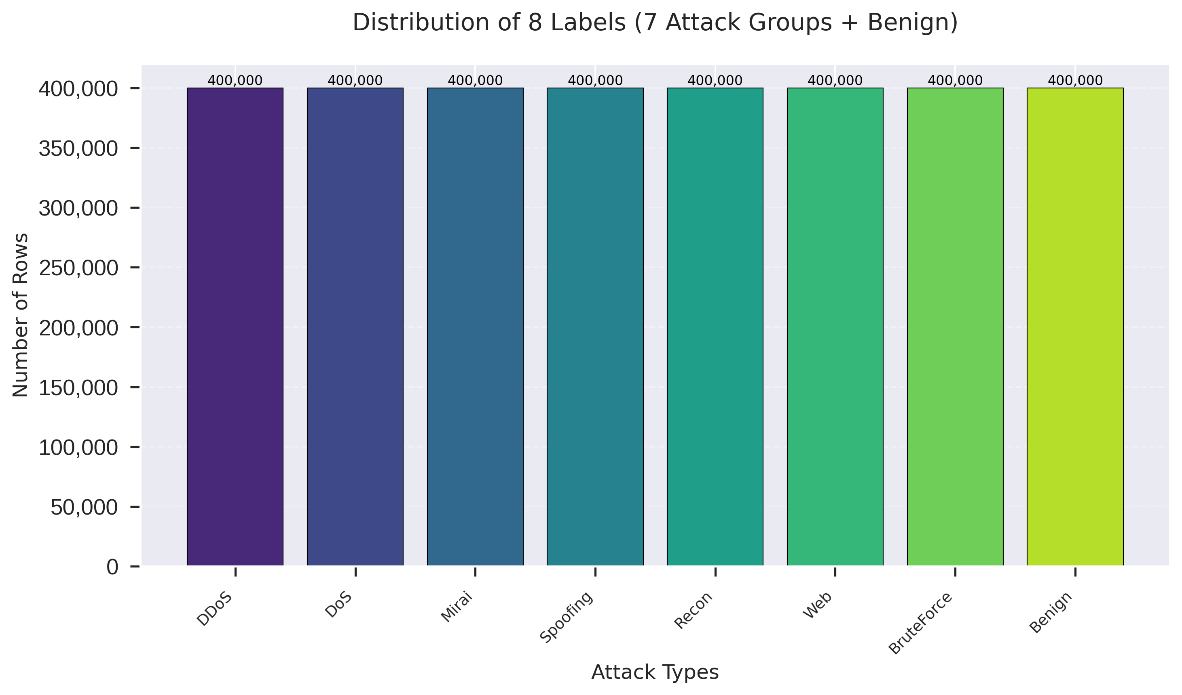
**2. Tập dữ liệu thực nghiệm**

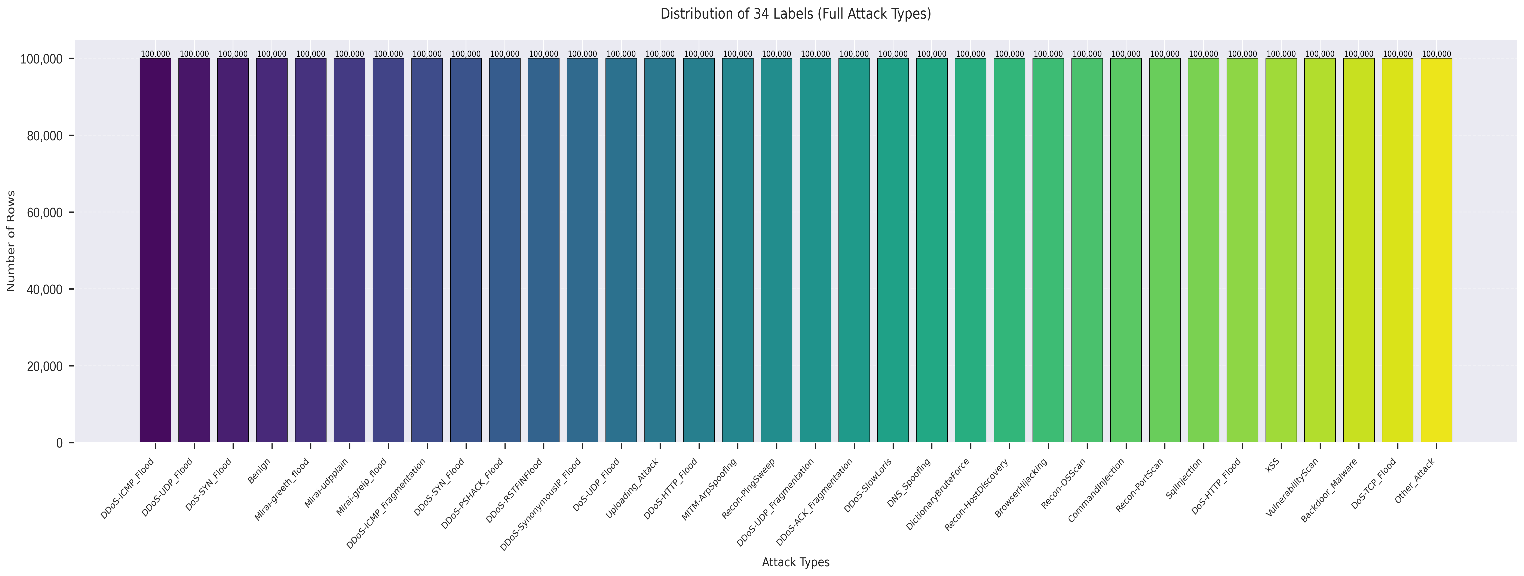
Nghiên cứu sử dụng tập dữ liệu CICIoT2023 [22], được phát triển bởi Trung tâm Nghiên cứu An ninh mạng Canada, chuyên về các kịch bản tấn công mạng trong môi trường Internet of Things (IoT). Tập dữ liệu này bao gồm lưu

lượng mạng từ các thiết bị IoT như camera, bộ định tuyến, và cảm biến, mô phỏng 33 loại tấn công khác nhau (DDoS, DoS, MITM, brute-force, reconnaissance, exploitation, spoofing). CICIoT2023 chứa tổng cộng 6.2 triệu mẫu dữ liệu, được chia thành tập huấn luyện (train) với 4.3 triệu mẫu (70%), tập xác thực (validation) với 1.2 triệu mẫu (20%), và tập kiểm tra (test) với 0.7 triệu mẫu (10%). Hình 3 minh họa phân bố các loại tấn công trong tập dữ liệu CICIoT2023, với tỷ lệ các lớp tấn công như DDoS (25,3%), DoS (18,9%), và MITM (16,0%)



**Hình 3** : Dữ liệu huấn luyện 2 nhãn



**Hình 4 :** Dữ liệu huấn luyện 8 nhãn

**Hình 5 :** Dữ luyện huấn luyện 34 nhãn

**B. Kịch bản huấn luyện**

**1. Huấn luyện trên 2 nhãn**

***Bảng 1 : Kịch bản huấn luyện trên 2 nhãn***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kịch bản | Backbone | Batch-size | Learning Rate | Epochs | Num classes |
| 1 | DCN-V2 | 4096 | 0.0001 | 100 | 2 |
| 2 | FT-Transformer | 4096 | 0.0001 | 100 | 2 |
| 3 | TabNet | 4096 | 0.0001 | 100 | 2 |
| 4 | TabR | 4096 | 0.0001 | 100 | 2 |
| 5 | Node-Gam | 4096 | 0.0001 | 100 | 2 |

Bảng 1 trình bày các kịch bản huấn luyện mô hình phân loại với 2 nhãn, sử dụng các backbone khác nhau gồm DCN-V2, FT-Transformer, TabNet, TabR và Node-Gam. Tất cả các kịch bản đều được huấn luyện với cùng cấu hình: batch-size là 4096, learning rate là 0.0001, số epoch là 100 và số lượng phân lớp (Num classes) là 2.

**2. Huấn luyện trên 8 nhãn**

***Bảng 2 : Kịch bản huấn luyện trên 8 nhãn***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kịch bản | Backbone | Batch-size | Learning Rate | Epochs | Num classes |
| 6 | DCN-V2 | 4096 | 0.0001 | 150 | 8 |
| 7 | FT-Transformer | 4096 | 0.0001 | 150 | 8 |
| 8 | TabNet | 4096 | 0.0001 | 150 | 8 |
| 9 | TabR | 4096 | 0.0001 | 150 | 8 |
| 10 | Node-Gam | 4096 | 0.0001 | 150 | 8 |

Bảng 2 mô tả các kịch bản huấn luyện mô hình phân loại với 8 nhãn, ứng dụng các kiến trúc backbone khác nhau như DCN-V2, FT-Transformer, TabNet, TabR và Node-Gam. Các mô hình đều được huấn luyện với cấu hình thống nhất gồm batch-size là 4096, learning rate là 0.0001, trong vòng 150 epoch, và số lớp đầu ra (Num classes) là 8.

**3. Huấn luyện trên 34 nhãn**

***Bảng 3 : Kịch bản huấn luyện trên 34 nhãn***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Kịch bản | Backbone | Batch-size | Learning Rate | Epochs | Num classes |
| 11 | DCN-V2 | 4096 | 0.0001 | 150 | 34 |
| 12 | FT-Transformer | 4096 | 0.0001 | 150 | 34 |
| 13 | TabNet | 4096 | 0.0001 | 150 | 34 |
| 14 | TabR | 4096 | 0.0001 | 150 | 34 |
| 15 | Node-Gam | 4096 | 0.0001 | 150 | 34 |

Bảng 3 liệt kê các kịch bản huấn luyện mô hình phân loại với số lượng nhãn lớn hơn, cụ thể là 34 nhãn. Các backbone được sử dụng bao gồm DCN-V2, FT-Transformer, TabNet, TabR và Node-Gam. Mỗi mô hình được huấn luyện với batch-size cố định là 4096, learning rate là 0.0001, số epoch là 150 và số lớp phân loại đầu ra là 34.

**C. Kết quả thực nghiệm trên 2 nhãn**

**1. Kết quả huấn luyện 2 nhãn**

*a. Độ đo loss của các kịch bản*

***Bảng 4 : Biểu đồ đo độ loss huấn luyện các mô hình trên 2 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 1 : DCN-v2** | **Kịch bản 2 : FT-Transformer** | **Kịch bản 3 : TabNet** |
|  |  |  |
| **Kịch bản 4 : TabR** | **Kịch bản 5 : Node-Gam** |  |

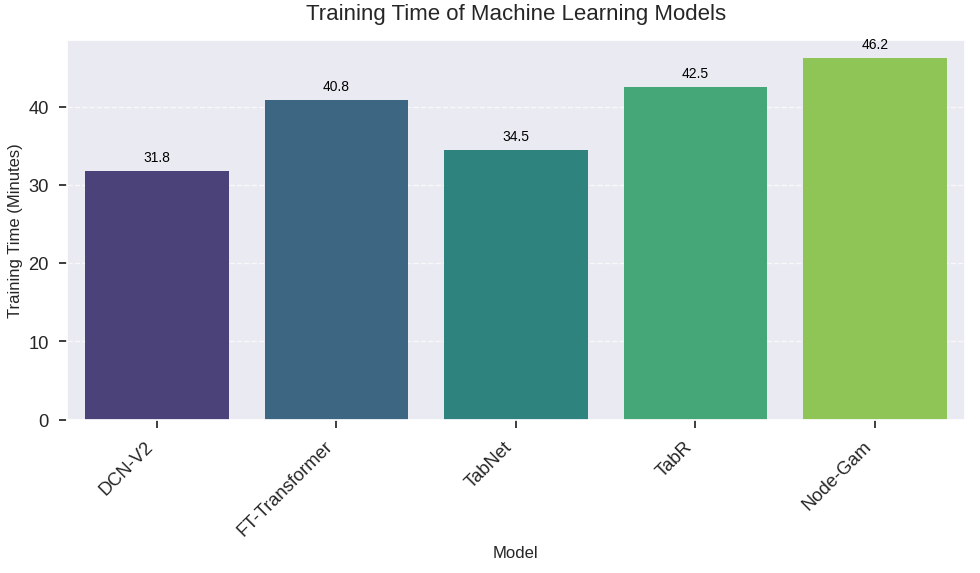
Biểu đồ trong Bảng 4 trình bày các giá trị độ đo loss của các kịch bản trong quá trình huấn luyện với 2 nhãn. Cụ thể, kịch bản 1 có training loss – validation loss lần lượt là 0.0018 – 0.0017, kịch bản 2 là 0.0017 – 0.0017, kịch bản 3 là 0.0017 – 0.0017, kịch bản 4 là 0.0018 – 0.0017 và kịch bản 5 là 0.0018 – 0.0018. Hầu hết các kịch bản đều có giá trị validation loss tương đương hoặc thấp hơn training loss, cho thấy mô hình học ổn định và không bị overfitting; trong đó, kịch bản 2 và 3 cho thấy sự cân bằng rõ rệt giữa hai giá trị loss.

*b. Độ đo Accuracy của các kịch bản*

***Bảng 5 : Biểu đồ đo độ accuracy huấn luyện các mô hình trên 2 nhãn***

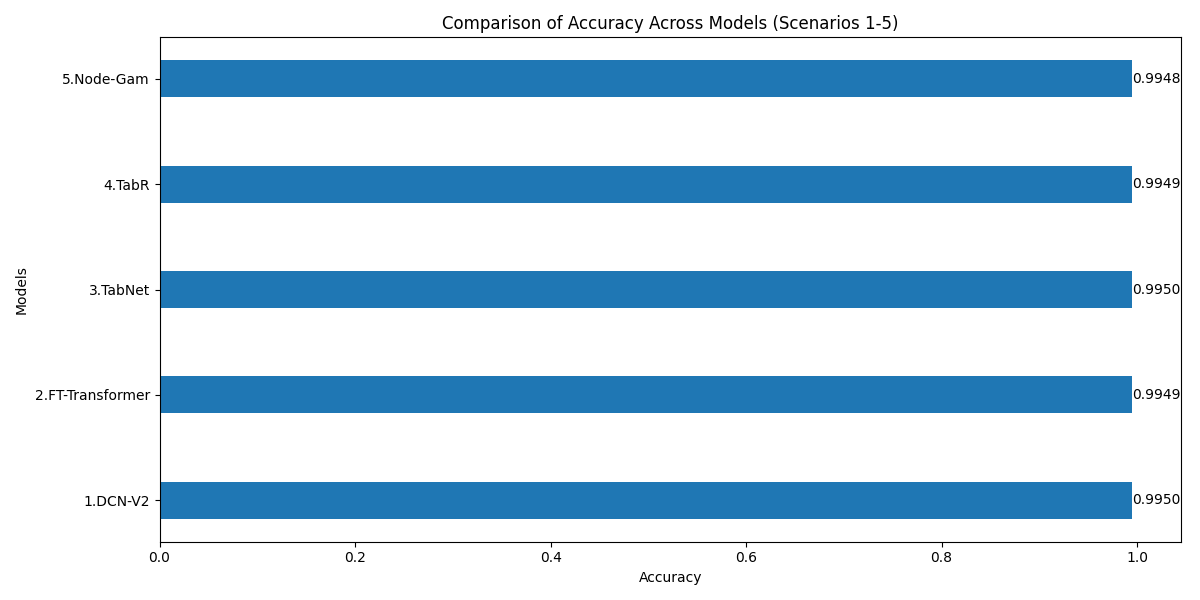
|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 1 : DCN-v2** | **Kịch bản 2 : FT-Transformer** | **Kịch bản 3 : TabNet** |
|  |  |  |
| **Kịch bản 4 : TabR** | **Kịch bản 5 : Node-Gam** |  |

Bảng 5 trình bày độ chính xác của các mô hình trên 2 nhãn. Cụ thể, kịch bản 1 (DCN-v2) có độ chính xác cao và ổn định với Train Accuracy và Val Accuracy đều đạt trung bình 99.47%. Các kịch bản còn lại dao động nhẹ quanh mức 99.46%–99.50%, cho thấy sự tương đương giữa các mô hình, trong đó kịch bản 3 (TabNet) có độ chính xác cao nhất 99.50%.

*****c. Thời gian huấn luyện*

**Hinh 6** : So sánh thời gian huấn luyện trên 2 nhãn

Hình 6 biểu diễn thời gian huấn luyện (tính bằng phút) của 5 mô hình máy học: DCNv2, FT-Transformer, TabNet, TabR và Node-Gam. Node-Gam có thời gian huấn luyện dài nhất (46,2 phút), tiếp theo là FT-Transformer (40,8 phút). DCNv2 có thời gian ngắn nhất (31,8 phút), trong khi TabR và TabNet lần lượt là 42,5 và 34,5 phút. Biểu đồ cho thấy sự khác biệt về hiệu suất huấn luyện giữa các mô hình, với Node-Gam chậm nhất và DCNv2 nhanh nhất.

**2. Kết quả kiểm thử 2 nhãn**

*a. Độ chính xác kiểm thử*

**Hình 7 :** So sánh kết quả kiểm thử trên 2 nhãn

Hình 7 thể hiện Xem biểu đồ độ chính xác của 5 mô hình trong kịch bản 1-5, mọi thứ khá đồng đều. DCN-V2 và TabNet nổi bật nhất với 0.9950, thật sự xuất sắc. FT-Transformer và TabR đạt 0.9949, chỉ kém một chút thôi. Node-Gam đứng cuối với 0.9948, nhưng vẫn rất ấn tượng.

*b. Giải thích kết quả bằng Shap*

***Bảng 6 : Giải thích kết quả kiểm thử bằng Shap trên 2 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 1 : DCN-v2** | **Kịch bản 2 : FT-Transformer** | **Kịch bản 3 : TabNet** |
|  |  |  |
| **Kịch bản 4 : TabR** | **Kịch bản 5 : Node-Gam** |  |

Bảng 6 cho thấy biểu đồ SHAP để đánh giá ảnh hưởng của 10 đặc trưng gốc đến dự đoán tấn công trên 5 mô hình: DCN-v2, FT-Transformer, TabNet, TabR và Node-Gam. Trong đó, "IAT", "Number", "urg\_count" là những đặc trưng có tác động lớn, nhưng mức độ ảnh hưởng khác nhau giữa các mô hình.

**D. Kết quả thực nghiệm trên 8 nhãn**

**1. Kết quả huấn luyện trên 8 nhãn**

*a. Độ đo loss của các kịch bản*

***Bảng 7 : Biểu đồ đo độ loss huấn luyện các mô hình trên 8 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 6 : DCN-v2** | **Kịch bản 7 : FT-Transformer** | **Kịch bản 8 : TabNet** |
|  |  |  |
| **Kịch bản 9 : TabR** | **Kịch bản 10 : Node-Gam** |  |

Bảng 7 hiển thị các biểu đồ học curve - loss của 5 mô hình (DCN-v2, FT-Transformer, TabNet, TabR, Node-Gam) trên 8 nhãn, cho thấy sự giảm dần của loss qua các epoch. Train Loss và Val Loss giảm mạnh ban đầu, sau đó ổn định, với TabNet đạt loss thấp nhất Avg Train Loss (0.0129) và Avg Val Loss (0.0127) dao động nhẹ, phản ánh sự phù hợp của mô hình, trong khi Node-Gam có loss cao nhất (0.0175 – 0.0163), cho thấy hiệu suất kém hơn so với các mô hình khác.

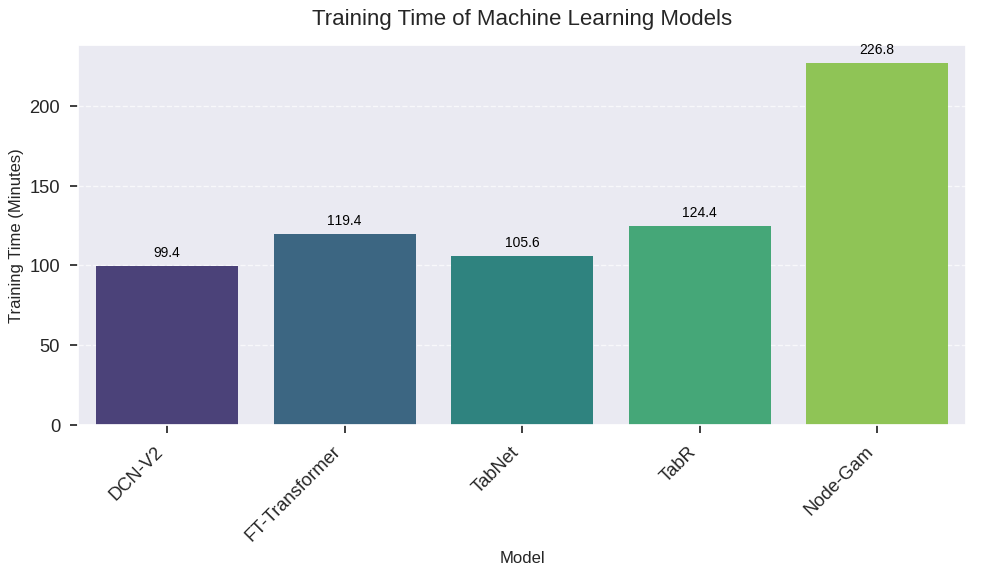
*b. Độ đo accuracy của các kịch bản*

***Bảng 8 : Biểu đồ đo độ accuracy huấn luyện các mô hình trên 8 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 7 : DCN-v2** | **Kịch bản 8 : FT-Transformer** | **Kịch bản 9 : TabNet** |
|  |  |  |
| **Kịch bản 10 : TabR** | **Kịch bản 11 : Node-Gam** |  |

Bảng 8 cho thấy hiệu suất Train Accuracy và Val Accuracy của 5 mô hình trên 8 nhãn. TabNet dẫn đầu với Train Acc 0.9295 và Val Acc 0.9313, thể hiện hiệu suất cao nhất. TabR có Val Acc bất thường 0.9947 (có thể là lỗi dữ liệu), trong khi Train Acc là 0.9218. DCN-V2 đạt Train Acc 0.9231 và Val Acc 0.9245, FT-Transformer đạt 0.9105 và 0.9117, còn Node-Gam thấp nhất với 0.9048 và 0.9124

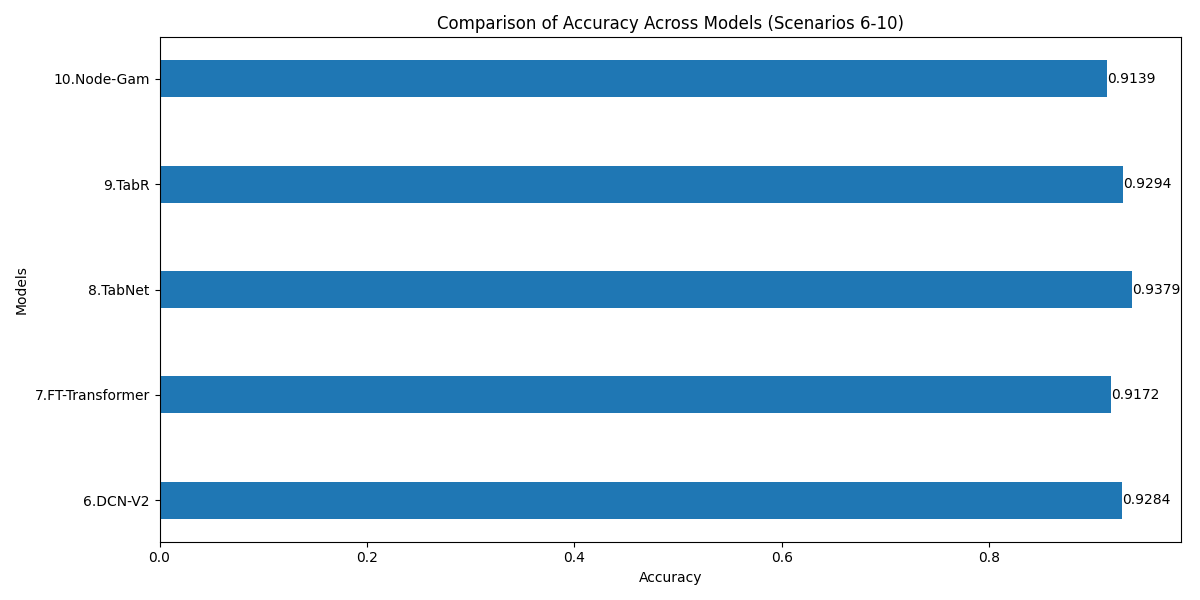
*c. Thời gian huấn luyện trên 8 nhãn*

****

**Hình 8:** So sánh thời gian huấn luyện các mô hình trên 8 nhãn

Hình 5 thể hiện DCN-V2 có thời gian ngắn nhất (99.38 phút), tiếp theo là TabNet (105.63 phút) và FT-Transformer (119.4 phút). TabR mất 124.4 phút, trong khi Node-Gam có thời gian dài nhất (226.58 phút), gấp đôi so với các mô hình khác. Kết quả này cho thấy DCN-V2 và TabNet huấn luyện nhanh nhất, trong khi Node-Gam là mô hình chậm nhất, cần tối ưu hóa để cải thiện hiệu quả.

**2. Kết quả kiểm thử trên 8 nhãn**

*a. Độ chính xác kiểm thử*

**Hình 9 :** So sánh kết quả kiểm thử trên 8 nhãn

Hình 9 cho thấy TabNet đạt độ chính xác cao nhất (0.9379), tiếp theo là TabR (0.9294) và DCN-V2 (0.9284). FT-Transformer đạt 0.9172, trong khi Node-Gam thấp nhất với 0.9139. Kết quả này khẳng định TabNet là mô hình hiệu quả nhất, trong khi Node-Gam cần cải thiện hiệu suất.

*b. Giải thích kết quả kiểm thử bằng Shap*

Bảng 9 : Giải thích kết quả kiểm thử bằng Shap trên 8 nhãn

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 11 : DCN-v2** | **Kịch bản 12 : FT-Transformer** | **Kịch bản 13 : TabNet** |
|  |  |  |
| **Kịch bản 14 : TabR** | **Kịch bản 15 : Node-Gam** |  |

Bảng 9 sử dụng SHAP Summary Plot để đánh giá tác động trung bình của 10 đặc trưng gốc trên dự đoán 8 nhãn (mapped, improved) với 5 mô hình: DCN-v2, FT-Transformer, TabNet, TabR và Node-Gam. Đặc trưng "IAT", "Tot size" và "Number" có ảnh hưởng lớn nhất ở tất cả các mô hình, với SHAP dao động từ 0.00 đến 0.20, trong đó TabR có phạm vi rộng nhất (đến 0.20). Các đặc trưng như "Protocol Type" và "Srate" có tác động thấp hơn, đặc biệt ở FT-Transformer và Node-Gam (SHAP chỉ đến 0.14).

**F. Kết quả thực nghiệm trên 34 nhãn**

**1.Kết quả huấn luyện trên 34 nhãn**

*a. Đọ đo loss các mô hình*

***Bảng 10 : Biểu đồ đo độ loss huấn luyện các mô hình trên 34 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 11 : DCN-v2** | **Kịch bản 12 : FT-Transformer** | **Kịch bản 13 : TabNet** |
|  |  |  |
| **Kịch bản 14 : TabR** | **Kịch bản 15 : Node-Gam** |  |

Bảng 10 Bảng dữ liệu cho thấy Train Loss và Val Loss của 5 mô hình trên 8 nhãn. FT-Transformer đạt Train Loss 0.0070 và Val Loss 0.0068, thấp nhất, cùng với Node-Gam (Val Loss 0.0068), cho thấy khả năng tối ưu tốt. TabNet có Train Loss 0.0076 và Val Loss 0.0069, trong khi DCN-V2 và TabR đều ghi nhận Train Loss 0.0081 và Val Loss 0.0072.

**b. Độ đo accuracy các mô hình**

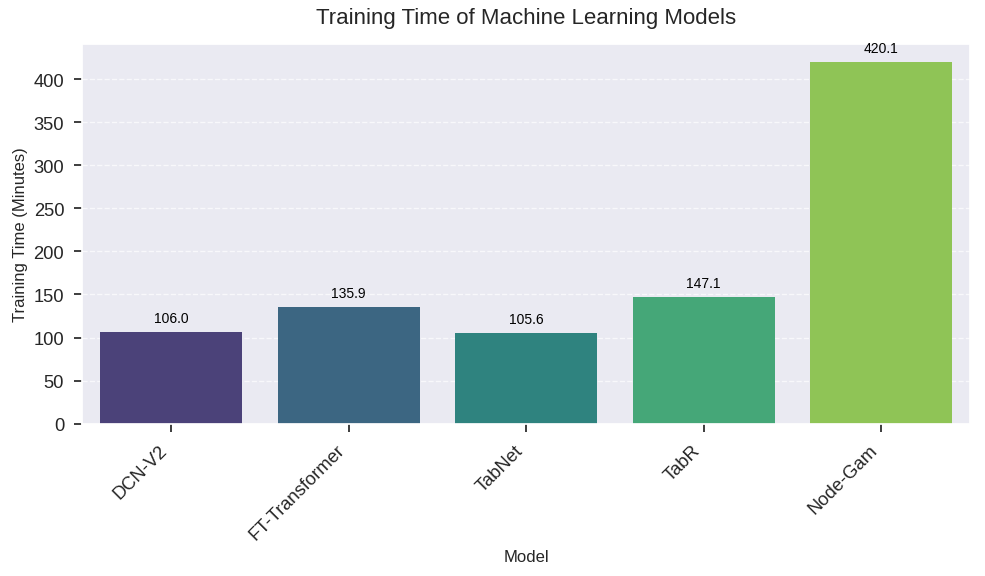
***Bảng 11 : Biểu đồ đo độ accuracy huấn luyện các mô hình trên 34 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 11 : DCN-v2** | **Kịch bản 12 : FT-Transformer** | **Kịch bản 13 : TabNet** |
|  |  |  |
| **Kịch bản 14 : TabR** | **Kịch bản 15 : Node-Gam** |  |

Bảng 11 chỉ rõ FT-Transformer dẫn đầu với Train Acc 0.9459 và Val Acc 0.9451, thể hiện hiệu suất cao nhất.

TabR đạt Train Acc 0.9322 và Val Acc 0.9414, trong khi DCN-V2 có Val Acc 0.9384 và Train Acc 0.9288. TabNet đạt 0.9295 và 0.9313, còn Node-Gam thấp nhất với 0.9266 và 0.9314. Kết quả khẳng định FT-Transformer là mô hình hiệu quả nhất.

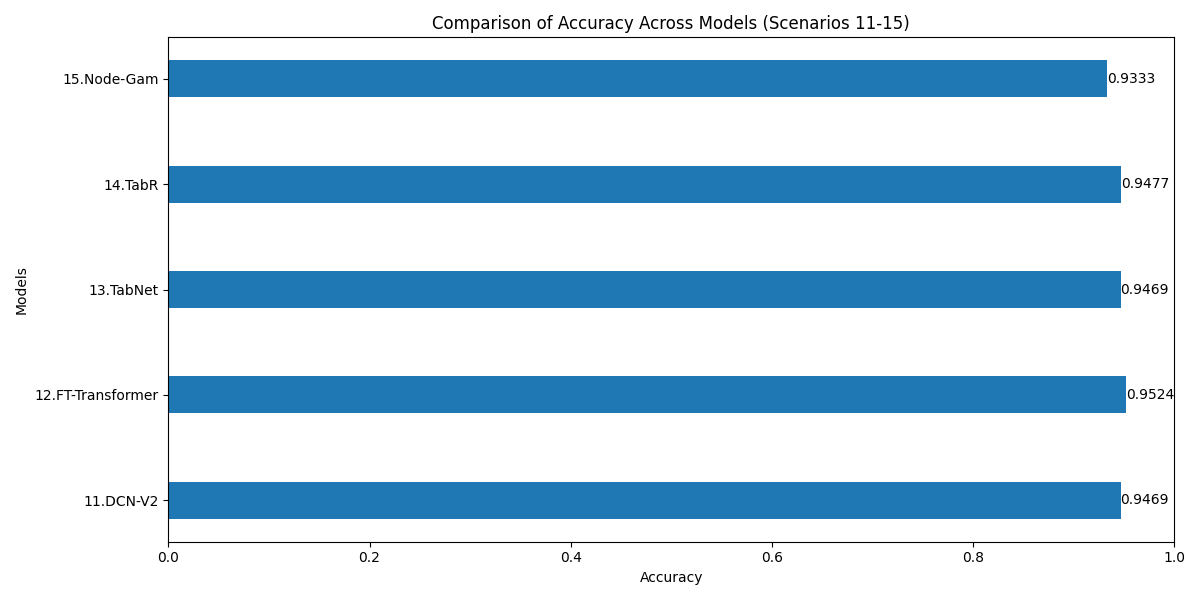
**c. Thời gian huấn luyện**

****

**Hình 10 :** So sánh thời gian huấn luyện các mô hình trên 34 nhãn

Hình 10 cho thấy cột thời gian huấn luyện (phút) của 5 mô hình trên 8 nhãn. DCN-V2 có thời gian ngắn nhất (106 phút), tiếp theo là TabNet (105.63 phút), và FT-Transformer (135.88 phút). TabR mất 147.13 phút, trong khi Node-Gam có thời gian dài nhất (420.08 phút), gấp nhiều lần so với các mô hình khác. Kết quả này cho thấy DCN-V2 và TabNet huấn luyện nhanh nhất, trong khi Node-Gam là mô hình chậm nhất.

**2. Kết quả kiểm thử trên 34 nhãn**

*a. Độ đo chính xác kiểm thử*

**Hình 11 :** So sánh kết quả kiểm thử trên 34 nhãn

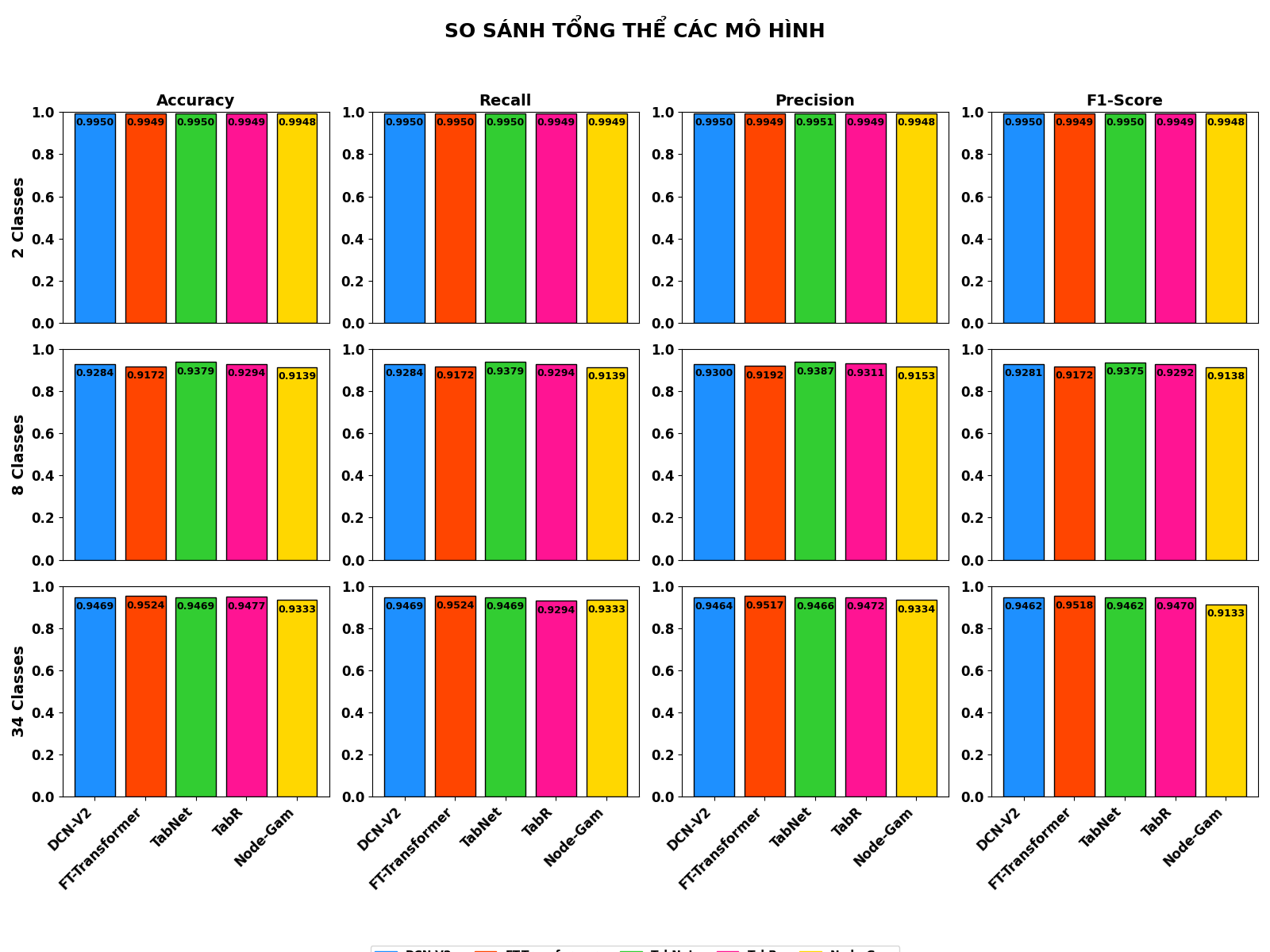
Hình 11 cho thấy độ chính xác (Accuracy) của 5 mô hình kiểm thử. FT-Transformer đạt cao nhất với 0.9524, tiếp theo là TabR (0.9477), DCN-V2 và TabNet cùng đạt 0.9469. Node-Gam có hiệu suất thấp nhất với 0.9333. Kết quả này khẳng định FT-Transformer là mô hình hiệu quả nhất.

*b. Giải tích kết quả kiểm thử bằng Shap*

***Bảng 12 : Giải thích kết quả kiểm thử bằng Shap trên 34 nhãn***

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Kịch bản 11 : DCN-v2** | **Kịch bản 12 : FT-Transformer** | **Kịch bản 13 : TabNet** |
|  |  |  |
| **Kịch bản 14 : TabR** | **Kịch bản 15 : Node-Gam** |  |

Bảng 12 thể hiện các SHAP Summary Plot để đánh giá tác động trung bình của 10 đặc trưng gốc trên dự đoán 34 nhãn với 5 mô hình: DCN-v2, FT-Transformer, TabNet, TabR và Node-Gam. Đặc trưng "Tot size", "IAT" và "Number" có ảnh hưởng lớn nhất, với SHAP dao động từ 0.00 đến 0.05 ở TabNet và TabR, trong khi các đặc trưng như "Weight" và "Duration" có tác động thấp hơn. FT-Transformer và Node-Gam cho thấy phạm vi SHAP rộng hơn (đến 0.04), trong khi TabR nổi bật với ảnh hưởng của "Telnet" và "TCP". Kết quả này cho thấy sự khác biệt trong cách các mô hình ưu tiên các đặc trưng.

**G. So sánh đánh giá kết quả các mô hình**

**Hình 12** : So sánh kết quả của các mô hình

Hình 12 cho thấy trong kịch bản 2 Classes, các mô hình đạt hiệu suất gần hoàn hảo (~0.995), TabNet và DCN-V2 nhỉnh hơn ở Precision. Với 8 Classes, TabNet dẫn đầu (0.9375–0.9387), Node-Gam thấp nhất (~0.9138). Ở 34 Classes, FT-Transformer vượt trội (0.9517–0.9524), Node-Gam kém nhất (~0.9133). TabNet và FT-Transformer nổi bật nhất, phù hợp với bài toán phức tạp. DCN-V2 và TabR ổn định, nhưng Node-Gam yếu hơn khi số lớp tăng.

**V. KẾT LUẬN**

Nghiên cứu đã ứng dụng thành công các mô hình học sâu tiên tiến (DCN-V2, FT-Transformer, TabNet, TabR, Node-Gam) để phát hiện tấn công mạng trên tập dữ liệu CICIoT2023, đạt hiệu suất vượt trội qua các kịch bản phân loại 2, 8 và 34 nhãn. TabNet dẫn đầu ở kịch bản 8 nhãn với độ chính xác 0.9379, trong khi FT-Transformer xuất sắc ở 34 nhãn (0.9524), thể hiện khả năng xử lý bài toán phức tạp. DCN-V2 và TabR duy trì hiệu suất ổn định, phù hợp cho các hệ thống IoT thời gian thực. Node-Gam có hiệu suất thấp hơn, đặc biệt khi số lớp tăng. Kết quả nhấn mạnh tiềm năng của học sâu trong nâng cao bảo mật IoT, giảm thiểu cảnh báo sai và phát hiện kịp thời các mối đe dọa. Nghiên cứu đề xuất tiếp tục tối ưu hóa mô hình, giảm chi phí tính toán và tích hợp vào các hệ thống thực tiễn để tăng cường an ninh mạng IoT.

**VI. TÀI LIỆU KHAM KHẢO**

[1] Statista, “IoT connections worldwide 2022-2033 | Statista.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[2] Kaspersky, “Kaspersky Detects 1.5B IoT Attacks This Year.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.pymnts.com/news/security-and-risk/2021/kaspersky-detects-iot-cyberattacks-double-last-year/

[3] M. Antonakakis *et al.*, “Understanding the Mirai Botnet”, Accessed: Apr. 09, 2025. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/antonakakis

[4] IBM Security., “Cost of a data breach 2024 | IBM.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.ibm.com/reports/data-breach

[5] I. , B. Y. , & C. Goodfellow, “Deep Learning.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.deeplearningbook.org/

[6] Z. Mo, Z. Zhang, and K. L. Tsui, “The variational kernel-based 1-D convolutional neural network for machinery fault diagnosis,” *IEEE Trans Instrum Meas*, vol. 70, 2021, doi: 10.1109/TIM.2021.3105252.

[7] Canadian Institute for Cybersecurity, “IoT Dataset 2023 | Datasets | Research | Canadian Institute for Cybersecurity | UNB.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.unb.ca/cic/datasets/iotdataset-2023.html

[8] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study,” *Journal of Information Security and Applications*, vol. 50, p. 102419, Feb. 2020, doi: 10.1016/J.JISA.2019.102419.

[9] J. Wang, T. S. Lembono, S. Kim, S. Calinon, S. Vijayakumar, and S. Tonneau, “Learning to Guide Online Multi-Contact Receding Horizon Planning,” *IEEE International Conference on Intelligent Robots and Systems*, vol. 2022-October, pp. 12942–12949, 2022, doi: 10.1109/IROS47612.2022.9981234.

[10] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, “Deep learning for cyber threat detection in IoT networks: A review,” *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 110–128, Jan. 2024, doi: 10.1016/J.IOTCPS.2023.09.003.

[11] E. C. P. Neto, S. Dadkhah, R. Ferreira, A. Zohourian, R. Lu, and A. A. Ghorbani, “CICIoT2023: A Real-Time Dataset and Benchmark for Large-Scale Attacks in IoT Environment,” *Sensors 2023, Vol. 23, Page 5941*, vol. 23, no. 13, p. 5941, Jun. 2023, doi: 10.3390/S23135941.

[12] “Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Apr. 09, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/9469914

[13] Q. A. Al-Haija and S. Zein-Sabatto, “An Efficient Deep-Learning-Based Detection and Classification System for Cyber-Attacks in IoT Communication Networks,” *Electronics 2020, Vol. 9, Page 2152*, vol. 9, no. 12, p. 2152, Dec. 2020, doi: 10.3390/ELECTRONICS9122152.

[14] S. I. Popoola, B. Adebisi, M. Hammoudeh, G. Gui, and H. Gacanin, “Hybrid Deep Learning for Botnet Attack Detection in the Internet-of-Things Networks,” *IEEE Internet Things J*, vol. 8, no. 6, pp. 4944–4956, Mar. 2021, doi: 10.1109/JIOT.2020.3034156.

[15] P. Sun *et al.*, “DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system,” *Security and Communication Networks*, vol. 2020, 2020, doi: 10.1155/2020/8890306.

[16] M. A. Haq and M. A. R. Khan, “DNNBoT: Deep Neural Network-Based Botnet Detection and Classification,” *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1729–1750, Nov. 2021, doi: 10.32604/CMC.2022.020938.

[17] Sophos, “The State of Ransomware 2024 – Sophos News.” Accessed: Apr. 09, 2025. [Online]. Available: https://news.sophos.com/en-us/2024/04/30/the-state-of-ransomware-2024/

[18] Check Point Software Technologies, “CYBER SECURIT Y REPORT,” 2024.

[19] S. Inc, “2024 SonicWall Cyber Threat Report,” 2024.

[20] I. Sainz, “Diseñar para divergencias y convergencias. Enfoques del DCG para los procesos de lectura por placer en la Red,” *Exploraciones, intercambios y relaciones entre el diseño y la tecnología*, pp. 57–79, 2019, doi: 10.16/CSS/JQUERY.DATATABLES.MIN.CSS.

[21] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset,” *Future Generation Computer Systems*, vol. 100, pp. 779–796, Nov. 2019, doi: 10.1016/J.FUTURE.2019.05.041.

[22] “IoT Dataset 2023 | Datasets | Research | Canadian Institute for Cybersecurity | UNB.” Accessed: Apr. 09, 2025. [Online]. Available: https://www.unb.ca/cic/datasets/iotdataset-2023.html