ISSN: 3126-2953
E-ISSN: 3126-3097

NHẬN DIỆN CÁC PHƯƠNG THỨC TẤN CÔNG MẠNG BẰNG PHƯƠNG PHÁP MÁY HỌC

IDENTIFICATION OF CYBER ATTACK METHODS USING MACHINE LEARNING TECHNIQUES

TRƯƠNG THÀNH KHANG[1,a], PHAN ANH CANG[1], PHAN THƯỢNG CANG[2]

[1]Trường Đại học Sư phạm Kỹ Thuật Vĩnh Long
[2]Trường Công nghệ thông tin & Truyền thông - Đại học Cần Thơ
[a]Tác giả liên hệ: khangtt@vlute.edu.vn

Phản biện (Reviewed):15/12/2025; Chấp nhận(Accepted):26/01/2026

TÓM TẮT

Nghiên cứu này tập trung vào bài toán phát hiện và phân loại tấn công mạng bằng các kỹ thuật học máy hiện đại, bao gồm CNN-LSTM, RNN và LSTM, trên tập dữ liệu NF-UQ-NIDS-v2. Hai kịch bản được triển khai: (1) huấn luyện trực tiếp trên dữ liệu gốc không xử lý mất cân bằng, và (2) áp dụng kỹ thuật cân bằng bằng trọng số lớp (class-weight). Kết quả thực nghiệm cho thấy trong kịch bản (1), các mô hình đạt độ chính xác cao với các lớp phổ biến nhưng gần như thất bại trong việc nhận diện các lớp hiếm. Ngược lại, ở kịch bản (2), phương pháp đánh trọng số giúp cải thiện khả năng phát hiện và phân loại toàn bộ 21 lớp, đặc biệt đối với các lớp thiểu số, mặc dù độ chính xác tổng thể của một số lớp đa số bị suy giảm. Nhìn chung, phương pháp cân bằng trọng số lớp đã chứng minh được hiệu quả trong việc giảm thiểu tác động của sự mất cân bằng dữ liệu, mở ra hướng nghiên cứu kết hợp với các kỹ thuật nâng cao như SMOTE biến thể nhằm tăng độ chính xác và tính khái quát hóa của mô hình trong thực tiễn.

Từ khóa: Phát hiện xâm nhập mạng, Máy học, Học sâu, CNN-LSTM, RNN, LSTM, Mất cân bằng dữ liệu, Trọng số lớp, NF-UQ-NIDS-v2, An ninh mạng

ABTRACT

This study focuses on the problem of network attack detection and classification using modern machine learning techniques, including CNN-LSTM, RNN, and LSTM, on the NF-UQ-NIDS-v2 dataset. Two experimental scenarios were conducted: (1) training directly on the original imbalanced dataset without any balancing technique, and (2) applying class-weight to address class imbalance. The experimental results indicate that in scenario (1), the models achieved high accuracy for majority classes but failed to recognize rare attack types. In contrast, scenario (2) showed that the class-weight approach significantly improved the detection and classification of all 21 classes, especially minority classes, although the accuracy of some majority classes decreased. Overall, the class-weight technique proved effective in mitigating the impact of data imbalance, suggesting future directions to integrate advanced methods such as SMOTE variants to further enhance accuracy and generalization capability in practical deployment.

Keywords: *Network Intrusion Detection, Machine Learning, Deep Learning, CNN-LSTM, RNN, LSTM, Class Imbalance, Class-weight, NF-UQ-NIDS-v2, Cybersecurity*

ISSN: 3126-2953
E-ISSN: 3126-3097

1. GIỚI THIỆU

1.1. Giới thiệu bài toán

Trong bối cảnh kỷ nguyên số hóa, Internet đã trở thành nền tảng quan trọng trong mọi khía cạnh của đời sống, từ cá nhân, doanh nghiệp đến chính phủ. Tuy nhiên, sự phát triển này cũng kéo theo sự gia tăng của các mối đe dọa an ninh mạng, với các cuộc tấn công ngày càng tinh vi và phức tạp. Những sự kiện nổi bật như WannaCry (2017), NotPetya (2017), và SolarWinds (2020) đã gây thiệt hại hàng tỷ đô la, ảnh hưởng đến hàng triệu người dùng toàn cầu và đặt ra thách thức lớn về an ninh mạng [1], [2]. Theo báo cáo mới nhất của IBM Security, chi phí trung bình của một vụ vi phạm dữ liệu đã tăng lên 4,88 triệu USD vào năm 2024, với dự báo tổng thiệt hại toàn cầu do tội phạm mạng có thể vượt quá 11 nghìn tỷ USD mỗi năm vào năm 2025 do sự gia tăng của các kỹ thuật tấn công sử dụng AI.

Các phương pháp bảo mật truyền thống như tường lửa hay phần mềm diệt virus dần trở nên kém hiệu quả trước sự thay đổi nhanh chóng của kỹ thuật tấn công [3]. Trong khi đó việc phát hiện sớm và ngăn chặn sớm chính là phương pháp hiệu quả nhất trong quá trình bảo vệ dữ liệu trước các cuộc tấn công mạng, mà các phương pháp máy học, đặc biệt là học sâu, nổi lên như một giải pháp tiềm năng nhờ khả năng tự động học hỏi từ dữ liệu lớn và phức tạp, từ đó có thể nhanh chóng phát hiện các mẫu tấn công mà các phương pháp cũ không thể nhận diện [4]. Nghiên cứu ứng dụng học sâu vào phát hiện tấn công mạng không chỉ mang ý nghĩa khoa học mà còn có giá trị thực tiễn cao, góp phần nâng cao khả năng bảo vệ hệ thống mạng trước các mối đe dọa hiện đại.

Đề tài nghiên cứu này sẽ tập trung vào đánh giá và so sánh hiệu quả của trong bài toán phân loại tấn công mạng đa lớp trên tập dữ liệu NF-UQ-NIDS-v2. Hai kịch bản huấn luyện được thiết kế nhằm phân tích tác động của sự mất cân bằng dữ liệu: huấn luyện trực tiếp trên dữ liệu gốc không áp dụng kỹ thuật cân bằng và áp dụng phương pháp đánh trọng số lớp (class-weight) để tăng mức độ ưu tiên học cho các lớp thiểu số. Thông qua phân tích chi tiết các chỉ số đánh giá và ma trận nhầm lẫn, nghiên cứu nhằm làm rõ ưu điểm, hạn chế cũng như sự đánh đổi giữa khả năng nhận diện các lớp hiếm và độ chính xác của các lớp đa số, từ đó đề xuất các hướng cải tiến phù hợp cho các nghiên cứu tiếp theo.

1.2. Các nghiên cứu liên quan

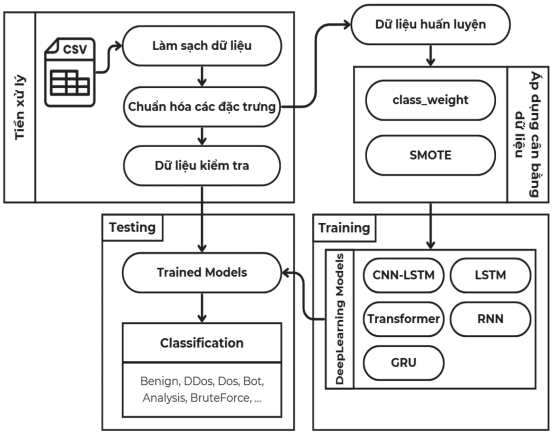
Máy học, đặc biệt là học sâu, đã chứng minh tiềm năng to lớn trong việc phát hiện và phân loại tấn công mạng. Các mô hình học sâu có khả năng học hỏi từ dữ liệu phức tạp và nhận diện các mẫu tấn công mà con người hoặc các hệ thống truyền thống khó phát hiện. Liu và cộng sự (2019) đã đề xuất sử dụng CNN và RNN để phân loại tải trọng mạng nhằm phát hiện các cuộc tấn công mạng trên tập dữ liệu DARPA1998. Kết quả cho thấy CNN đạt độ chính xác cao nhất trong các chỉ số đánh giá, đặc biệt trong việc nhận diện các mẫu tấn công phức tạp. Nghiên cứu này nhấn mạnh khả năng của CNN trong việc xử lý dữ liệu không gian và RNN trong việc phân tích dữ liệu tuần tự [1]. Hagar và Gawali (2022) đã sử dụng thuật toán Random Forest để chọn lọc 19 trường dữ liệu quan trọng từ 84 trường trong bộ dữ liệu CSE-CIC-IDS2018, nhằm giảm kích thước dữ liệu và

ISSN: 3126-2953
E-ISSN: 3126-3097

tăng hiệu quả huấn luyện. Kết hợp với các mô hình CNN và LSTM, nghiên cứu này đạt độ chính xác cao trong việc nhận diện 14 loại tấn công mạng, với thời gian huấn luyện được rút ngắn còn 7,56 phút nhờ sử dụng Apache Spark [5]. Long và cộng sự (2024) đã phát triển một hệ thống IDS dựa trên Transformer cho môi trường đám mây, đạt độ chính xác trên 90% trong việc phát hiện 7 loại tấn công mạng. Mô hình này có hiệu suất tương đương với mô hình CNN-LSTM, nhưng vượt trội về khả năng song song hóa và giảm thời gian huấn luyện [6]. Wang và cộng sự (2023) đã đề xuất một phương pháp phát hiện xâm nhập mạng dựa trên học sâu, kết hợp CNN và LSTM nhằm khai thác đồng thời đặc trưng không gian và đặc trưng chuỗi thời gian của lưu lượng mạng. Bộ dữ liệu CSE-CIC-IDS2018, chứa nhiều loại tấn công mạng hiện đại, được sử dụng cho huấn luyện và đánh giá. Trong giai đoạn tiền xử lý, dữ liệu được làm sạch, chuẩn hóa, mã hóa nhãn và thực hiện chọn lọc đặc trưng nhằm giảm nhiễu và cải thiện tốc độ huấn luyện. Mô hình CNN đảm nhận vai trò trích xuất đặc trưng cục bộ từ chuỗi dữ liệu, sau đó LSTM tiếp nhận và học các phụ thuộc theo thời gian để phân loại cuối cùng. Thử nghiệm cho thấy mô hình CNN–LSTM đạt độ chính xác 99,4%, F1-score 99,3% và tỷ lệ phát hiện tấn công (Recall) cao hơn các mô hình đối chứng như CNN đơn, LSTM đơn và một số phương pháp học máy truyền thống. Kết quả này khẳng định hiệu quả của kiến trúc lai CNN–LSTM trong phát hiện và phân loại chính xác nhiều loại tấn công mạng phức tạp [7]. Ashiku và Dagli (2021) đã nghiên cứu hệ thống phát hiện xâm nhập mạng dựa trên học sâu, so sánh hiệu năng của ba mô hình: MLP, CNN và LSTM. Bộ dữ liệu CIC-IDS2017, chứa nhiều dạng tấn công mạng đa dạng, được sử dụng cho huấn luyện và đánh giá. Quá trình tiền xử lý bao gồm làm sạch dữ liệu, chuẩn hóa, và mã hóa nhãn để chuẩn bị đầu vào cho mạng nơ-ron. MLP được dùng làm mô hình cơ sở, CNN đảm nhiệm trích xuất đặc trưng không gian, trong khi LSTM khai thác đặc trưng phụ thuộc theo thời gian. Kết quả thực nghiệm cho thấy CNN đạt độ chính xác cao nhất (~99,6%), vượt trội hơn LSTM (~99,2%) và MLP (~98,7%). Ngoài ra, CNN cũng cho F1-score và Recall cao nhất, chứng tỏ khả năng cân bằng tốt giữa phát hiện tấn công và giảm tỷ lệ báo động giả. Nghiên cứu khẳng định tiềm năng của CNN trong phát hiện xâm nhập mạng khi xử lý dữ liệu lưu lượng phức tạp và đa dạng [8].

2. PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mô hình đề xuất



Hình 1. Quy trình nghiên cứu và mô hình sử dụng

Mô hình tổng quát được đề xuất trong nghiên cứu, mô tả toàn bộ quy trình từ xử lý dữ liệu đầu vào đến huấn luyện và đánh giá mô hình học sâu trên tập dữ liệu NF-UQ-NIDS-v2. Quy trình bắt đầu từ việc đọc dữ liệu gốc ở định dạng CSV, sau đó tiến hành làm sạch dữ liệu nhằm loại bỏ các giá trị nhiễu, thiếu hoặc không hợp lệ. Tiếp theo, dữ liệu được chuẩn hóa đặc trưng để đảm bảo sự đồng nhất về thang

ISSN: 3126-2953
E-ISSN: 3126-3097

đo và cải thiện hiệu quả huấn luyện. Ở giai đoạn huấn luyện, dữ liệu được áp dụng các kỹ thuật cân bằng dữ liệu như class_weight hoặc SMOTE nhằm giảm thiểu ảnh hưởng của mất cân bằng lớp, giúp mô hình chú trọng hơn đến các lớp thiểu số. Sau đó, dữ liệu huấn luyện được đưa vào các mô hình học sâu gồm CNN-LSTM, LSTM và RNN để tiến hành huấn luyện và tối ưu tham số cuối cùng là so sánh kết quả với mô hình không áp dụng các biện pháp cân bằng để tìm ra mô hình tối ưu cho toàn bộ bài toán. Khi huấn luyện hoàn tất, các mô hình đã học được sử dụng để dự đoán và phân loại trên tập dữ liệu kiểm thử, với các nhãn đầu ra tương ứng với các loại tấn công mạng và lưu lượng bình thường (ví dụ: Benign, DDoS, DoS, Bot, Analysis, …).

2.2. Kịch bản huấn luyện

Bảng 1. Các kịch bản huấn luyện

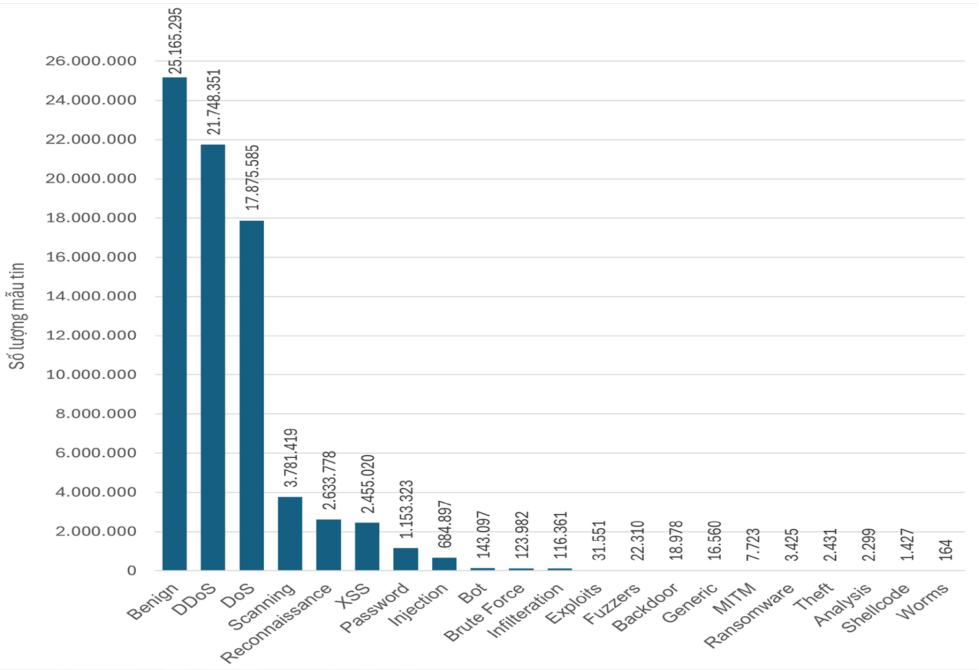
Kịch bản	Mô hình sử dụng	Kỹ thuật cân bằng	batch size	num classes	learning rate	max epoch
1	CNN-LSTM	Không áp dụng	512	(43, 21)	0,0001	500
	RNN					
	LSTM					
2	CNN-LSTM	class_weight	512	(43, 21)	0,0001	500
	RNN					
	LSTM					

Nghiên cứu này thực hiện 2 kịch bản: Với kịch bản 1 tập dữ liệu đầu vào sẽ được giữ nguyên không áp dụng bất kỳ phương pháp cân bằng dữ liệu nào. Tập dữ liệu trên kịch bản 2 sẽ được áp dụng phương pháp cân bằng đánh trọng số lớp (class_weight).

2.3. Mô tả tập dữ liệu

NF-UQ-NIDS-v2: Một bộ dữ liệu toàn diện, hợp nhất tất cả các bộ dữ liệu: NF-BoT-IoT-v2, NF-CSE-CIC-IDS2018-v2, NF-ToN-IoT-v2, NF-UNSW-NB15-v2. Tập dữ liệu NF-UQ-NIDS-v2 là một bộ dữ liệu mở được cung cấp trên nền tảng Kaggle, bao gồm phiên bản thứ hai của các tập dữ liệu Netflow được mở rộng với 43 đặc trưng NetFlow. Tập dữ liệu này được thiết kế để hỗ trợ nghiên cứu và phát triển các hệ thống phát hiện xâm nhập mạng (Network Intrusion Detection Systems - NIDS). Với cấu trúc dữ liệu phong phú, NF-UQ-NIDS-v2 cung cấp thông tin chi tiết về lưu lượng mạng, bao gồm các đặc trưng như lưu lượng byte, cờ TCP, và thời gian luồng, giúp mô phỏng các kịch bản tấn công mạng thực tế. Bộ dữ liệu này là một nguồn tài nguyên quan trọng để huấn luyện và đánh giá các mô hình học máy, đặc biệt trong việc phân loại các loại tấn công như DDoS, DoS, và các mối đe dọa hiếm gặp khác, nhờ vào tính đa dạng và quy mô lớn của nó.

ISSN: 3126-2953
E-ISSN: 3126-3097



Hình 2. Phân bố mẫu tin giữa các lớp

2.4. Tiền xử lý

Tập dữ liệu sẽ được đọc vào từ định dạng CSV, sau đó sẽ được làm sạch bằng cách loại bỏ các mẫu tin bị lỗi hoặc không có giá trị. Hai cột Label và Dataset bị loại bỏ còn lại 43 cột thông tin và cột Attack chứa nhãn phân loại, 43 cột này được chuyển đổi sang kiểu DoubleType, trong khi các cột chuỗi sẽ được mã hóa thành số bằng StringIndexer. Các đặc trưng sau khi được mã hóa sẽ được tập hợp bằng VectorAssembler và được chuẩn hóa bằng StandardScaler về cùng 1 thang đo. Riêng cột nhãn sẽ được mã hóa thành các số nguyên từ 0-20. Cuối cùng tập dữ liệu sau xử lý được chia thành 3 phần: 70% cho tập huấn luyện và tập đánh giá và tập kiểm tra mỗi tập được chia 15% được lưu dưới dạng file .npz nén.

2.4.1. Xử lý mất cân bằng

Tập dữ liệu sẽ được áp dụng kỹ thuật đánh trọng số lớp, trọng số của mỗi lớp được tính toán dựa trên công thức tỉ lệ nghịch với số lượng mẫu của lớp đó, sao cho các lớp ít mẫu sẽ có trọng số cao hơn. Điều này giúp mô hình tăng mức độ ưu tiên trong việc học các lớp thiểu số, đồng thời giảm ảnh hưởng của các lớp chiếm đa số mà không cần thay đổi phân phối dữ liệu gốc.

Công thức tính trọng số lớp:

$$\omega_c = \frac{N}{count_c * C}$$

Trong đó: N là tổng số mẫu trong tập dữ liệu. counte là số lượng mẫu thuộc lớp c và c là tổng số lớp trong tập dữ liệu.

Bảng 2. Bảng phân phối trọng số

Nhãn	Tên lớp	Số lượng mẫu	Trọng số
0	Benign	25.165.295	0,1438
1	DDoS	21.748.351	0,1664

ISSN: 3126-2953
E-ISSN: 3126-3097

Nhãn	Tên lớp	Số lượng mẫu	Trọng số
2	DoS	17.875.585	0,2024
3	scanning	3.781.419	0,9569
4	Reconnaissance	2.633.778	1,3739
5	xss	2.455.020	1,4739
6	password	1.153.323	3,1374
7	injection	684.897	5,2832
8	Bot	143.097	25,2869
9	Brute Force	123.982	29,1855
10	Infilteration	116.361	31,097
11	Exploits	31.551	114,6865
12	Fuzzers	22.310	162,1907
13	Backdoor	18.978	190,6668
14	Generic	16.560	218,5069
15	mitm	7.723	468,5323
16	ransomware	3.425	1.056,489
17	Theft	2.431	1.488,472
18	Analysis	2.299	1.573,934
19	Shellcode	1.427	2.535,722
20	Worms	164	22.063,87

Trọng số lớp giúp cải thiện recall và F1-score của các lớp hiếm, làm tăng macro-average F1 và khả năng bao phủ lớp. Đây là yếu tố quan trọng trong bài toán phát hiện tấn công mạng, vì nhiều loại tấn công nguy hiểm (như Worms, Ransomware) vốn có tần suất xuất hiện thấp. Bên cạnh đó, việc gán trọng số rất lớn cho các lớp cực hiếm có thể gây dao động gradient và làm suy giảm hiệu năng của các lớp đa số, đặc biệt là lớp benign.

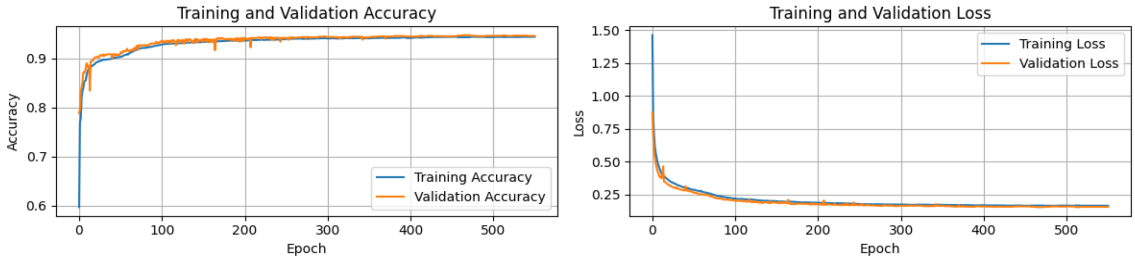
3. KẾT QUẢ NGHIÊN CỨU

3.1. Kịch bản 1

Trong kịch bản này, mô hình được huấn luyện trực tiếp trên tập dữ liệu gốc mà không áp dụng bất kỳ kỹ thuật xử lý mất cân bằng nào. Các đặc trưng được chuẩn hóa, và nhãn được mã hóa thành số nguyên, nhưng phân phối lớp giữ nguyên như dữ liệu gốc.

ISSN: 3126-2953
E-ISSN: 3126-3097

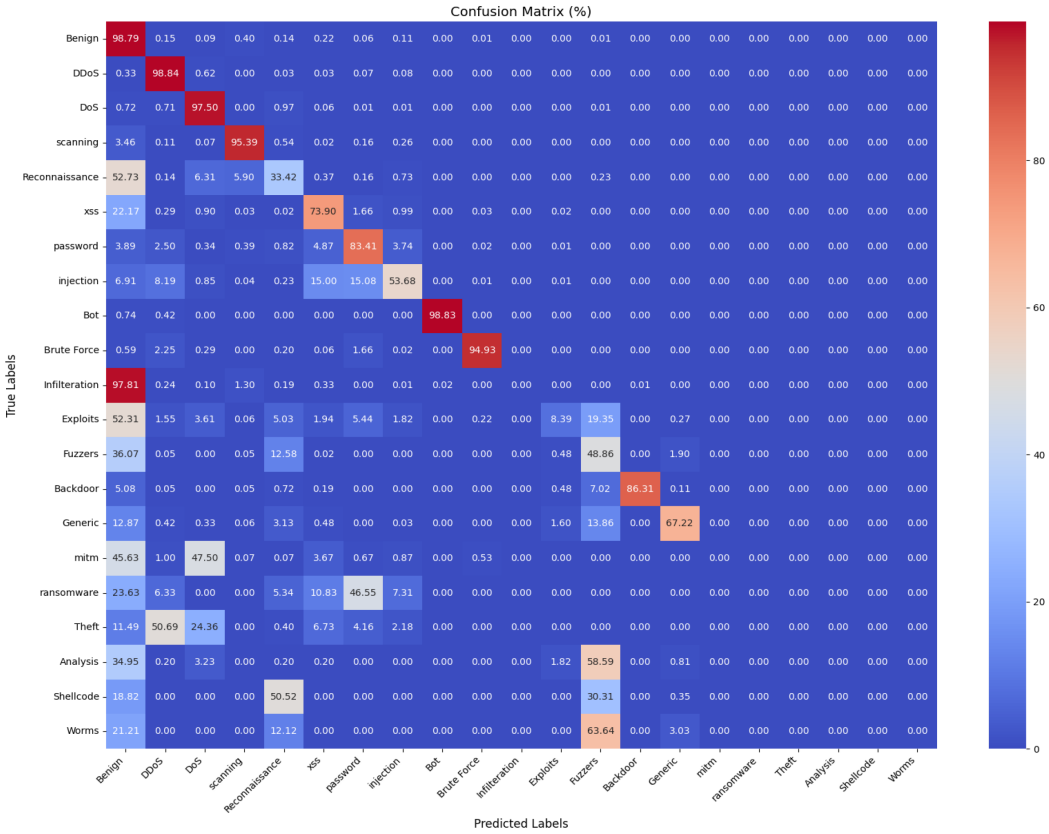
3.1.1. Mô hình CNN-LSTM



(a) Accuracy (b) Loss

Hình 3. Chỉ số huấn luyện mô hình CNN-LSTM

Đường cong độ chính xác trên tập huấn luyện và tập kiểm tra đều rất mượt mà và tăng nhanh, sau đó ổn định ở mức cao. Đường cong độ mất mát giảm dần và ổn định ở mức thấp. Sự ổn định này không phải là dấu hiệu của một mô hình hiệu quả toàn diện, mà là do mô hình đang tập trung vào việc học các đặc trưng của các lớp chiếm đa số, giúp nó giảm hàm mất mát nhanh chóng và không gặp phải sự dao động lớn.



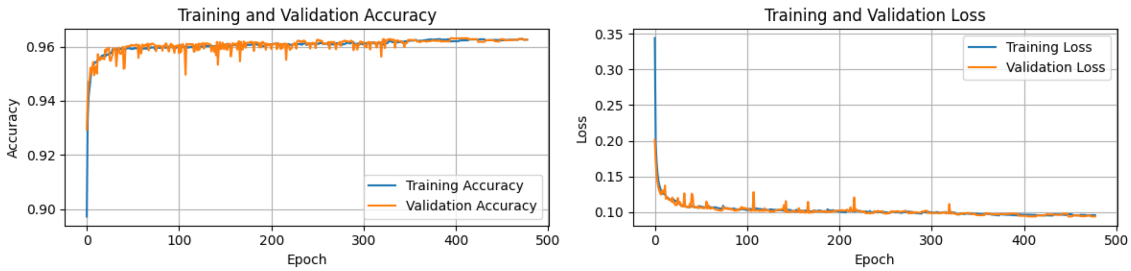
Hình 4. Ma trận nhầm lẫn mô hình CNN-LSTM

Ma trận nhầm lẫn cho thấy rõ sự mất cân bằng. Các giá trị trên đường chéo chính của các lớp chiếm đa số rất cao (benign 98,79%, ddos 98,84%, dos 97,50%...), cho thấy mô hình phân loại các lớp này rất tốt. Ngược lại, các hàng tương ứng với các lớp cực kỳ hiếm gần như bằng 0 (infilteration, mitm, ransomware, …), chứng tỏ mô hình

ISSN: 3126-2953
E-ISSN: 3126-3097

không dự đoán đúng các mẫu của chúng. Thay vào đó, các mẫu này thường bị phân loại nhầm thành các lớp đa số, đặc biệt là Benign, DDoS, và scanning, đặc biệt lớp infilteration bị nhầm lẫn sang lớp benign tới 97.81%

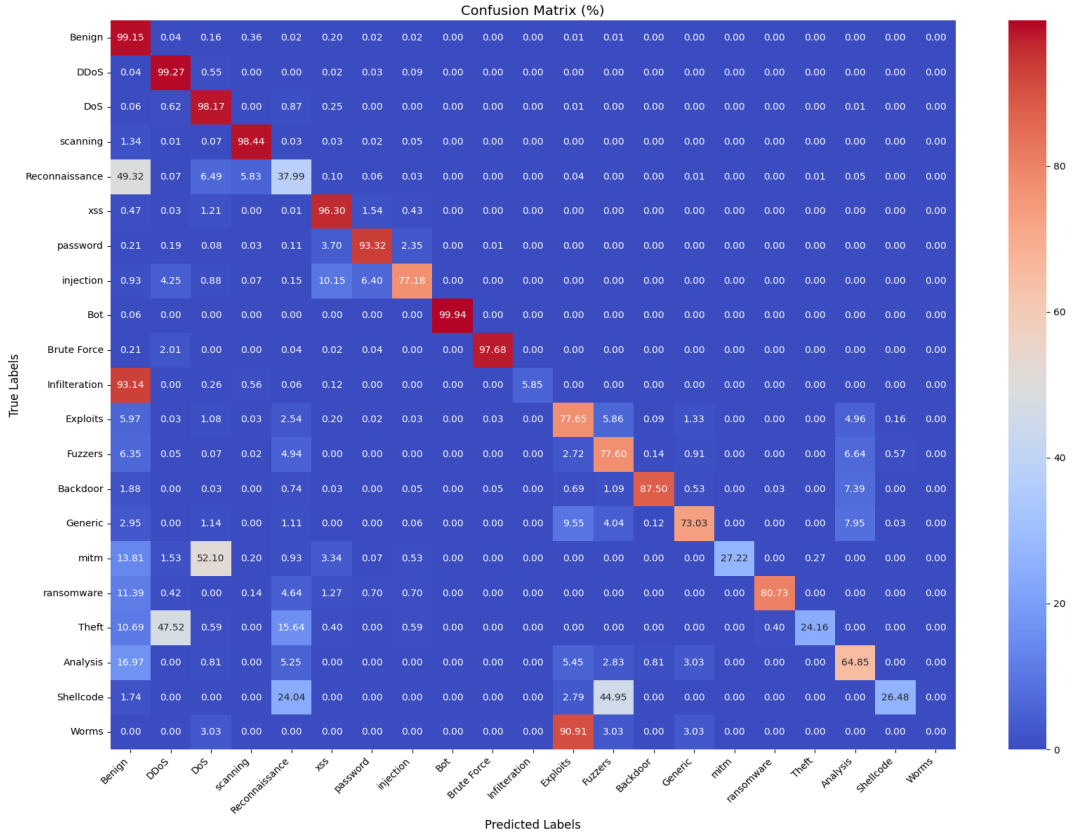
3.1.2. Mô hình RNN



(a) Accuracy (b) Loss

Hình 5. Chỉ số huấn luyện mô hình RNN

Biểu đồ lịch sử huấn luyện của RNN cho thấy một quá trình học tập ổn định. Độ chính xác trên cả tập huấn luyện và tập kiểm tra đều tăng dần và hội tụ ở mức cao. Tương tự, độ mất mát giảm dần và ổn định. Cùng với các dao động nhỏ trong quá trình huấn luyện cho thấy mô hình không bị overfitting.



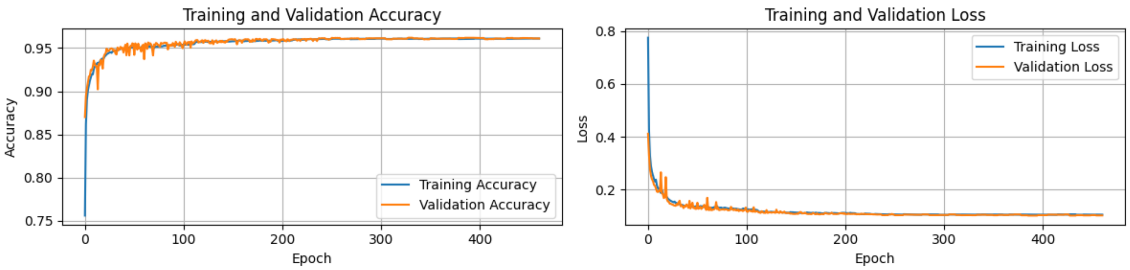
Hình 6. Ma trận nhầm lẫn mô hình RNN

Màu sắc đường chéo chính trên ma trận nhầm lẫn của RNN cho thấy khả năng phân loại tốt các lớp chiếm đa số như Benign, DDoS, và DoS. Tuy nhiên, một số lớp thiểu số có số lượng mẫu lớn hơn như Reconnaissance, Infilteration, Theft,

ISSN: 3126-2953
E-ISSN: 3126-3097

Analysis, và Shellcode vẫn bị phân loại sai rất nhiều, còn lớp Worms hầu như bị nhầm lẫn hoàn toàn sang lớp Exploits. Tại kịch bản này thì mô hình RNN cho thấy sự vượt trội so với các mô hình khác về khả năng tổng quát hóa phân loại được đa số các lớp, mặc dù chưa được tốt ở các lớp thiểu số (đặc biệt là chưa nhận diện được lớp Worms).

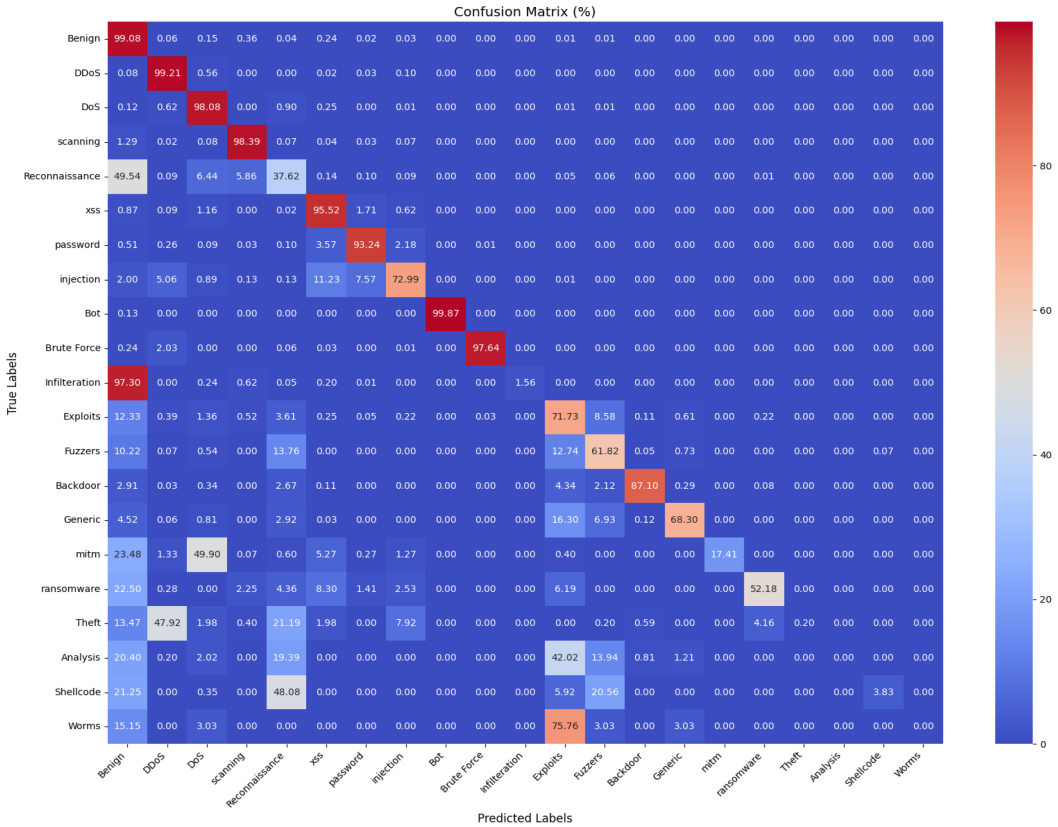
3.1.3. Mô hình LSTM



(a) Accuracy (b) Loss

Hình 7. Chỉ số huấn luyện mô hình LSTM

Biểu đồ lịch sử huấn luyện của LSTM cho thấy quá trình học tập diễn ra ổn định và mượt mà. Độ chính xác có dao động nhẹ lúc đầu sau đó ổn định và tăng dần đều và Độ mất mát cũng như vậy ổn định và giảm dần về mức thấp. Nhìn tổng quát có vẻ mô hình đang được huấn luyện tốt nhưng thật ra với sự mất cân bằng nghiêm trọng của tập dữ liệu thì đây là dấu hiệu của việc mô hình đang quá tập trung vào các lớp đa số mà bỏ qua các lớp thiểu số.

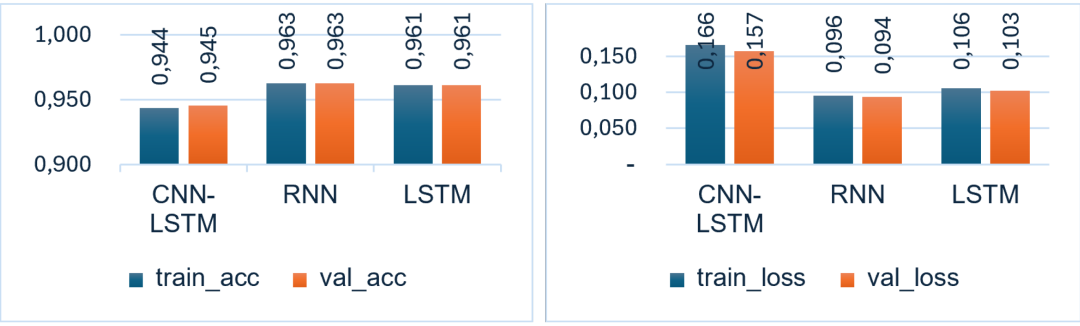


Hình 8. Ma trận nhầm lẫn mô hình LSTM

ISSN: 3126-2953
E-ISSN: 3126-3097

Ma trận nhầm lẫn cho thấy rõ ràng vấn đề mất cân bằng dữ liệu. Mô hình phân loại các lớp đa số với độ chính xác rất cao, thể hiện qua các giá trị trên đường chéo chính của các lớp này. Ngược lại, các mẫu từ các lớp thiểu số bị phân loại nhầm sang các lớp khác rất nhiều. Đặc biệt, các mẫu từ các lớp cực kỳ hiếm như Infilteration, ransomware, Theft, Analysis, Shellcode và Worms gần như hoàn toàn bị dự đoán sai, thường rơi vào lớp Benign hoặc các lớp đa số khác.

3.1.4. Tổng quan kết quả kịch bản 1



(a) Accuracy (b) Loss

Hình 9. So sánh Accuracy và Loss của các mô hình trong kịch bản 1

Bảng 3. So sánh hiệu năng các mô hình theo Macro metrics trong kịch bản 1

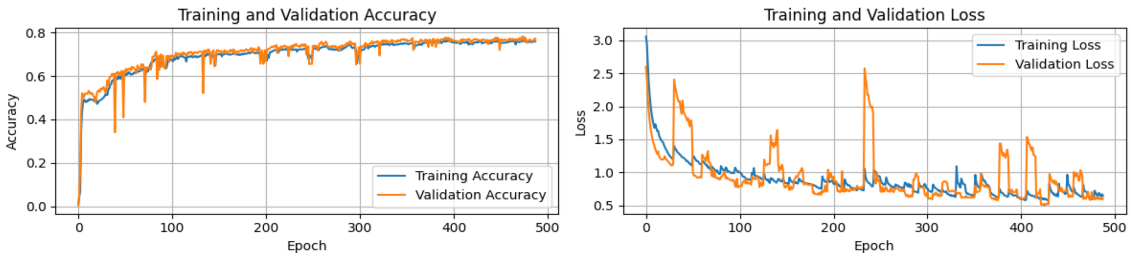
Mô hình	Accuracy	Macro-Precision	Macro-Recall	Macro-F1	Weighted-F1
CNN-LSTM	0,94	55,35	49,5	51,02	93,74
RNN	0,96	80,99	68,69	70,09	95,75
LSTM	0,96	76,14	59,8	62,06	95,57

Tất cả các mô hình đều cho thấy quá trình huấn luyện ổn định trên tập dữ liệu gốc với độ chính xác tăng và hàm mất mát giảm dần. Đáng chú ý hơn là sự chênh lệch lớn giữa Macro-F1 và Weighted-F1 ở cả ba mô hình cho thấy độ chính xác tổng thể cao chủ yếu đến từ các lớp đa số, trong khi hiệu năng trên các lớp thiểu số vẫn còn hạn chế. Tuy nhiên, điều này chủ yếu phản ánh việc mô hình tập trung vào các lớp đa số và chưa xử lý hiệu quả các lớp thiểu số. Về hiệu suất tổng thể, ba mô hình đạt kết quả tương đối tương đồng, trong đó mô hình RNN thể hiện ưu thế hơn về khả năng nhận diện nhiều lớp dựa trên ma trận nhầm lẫn, dù các lớp thiểu số vẫn còn bị nhầm lẫn đáng kể. Trong khi đó, mô hình CNN–LSTM cho thấy hiệu năng thấp hơn đáng kể so với hai mô hình còn lại, đặc biệt khi xét theo các chỉ số Macro-Precision, Macro-Recall và Macro-F1, cho thấy mô hình này chưa thực sự phù hợp với kịch bản hiện tại. Nhìn chung, các mô hình phân loại tốt các dạng tấn công phổ biến nhưng còn hạn chế đối với các tấn công hiếm do sự mất cân bằng dữ liệu, đặt ra nhu cầu nghiên cứu các phương pháp xử lý mất cân bằng và mô hình học nâng cao trong các nghiên cứu tiếp theo.

ISSN: 3126-2953
E-ISSN: 3126-3097

3.2. Kịch bản 2

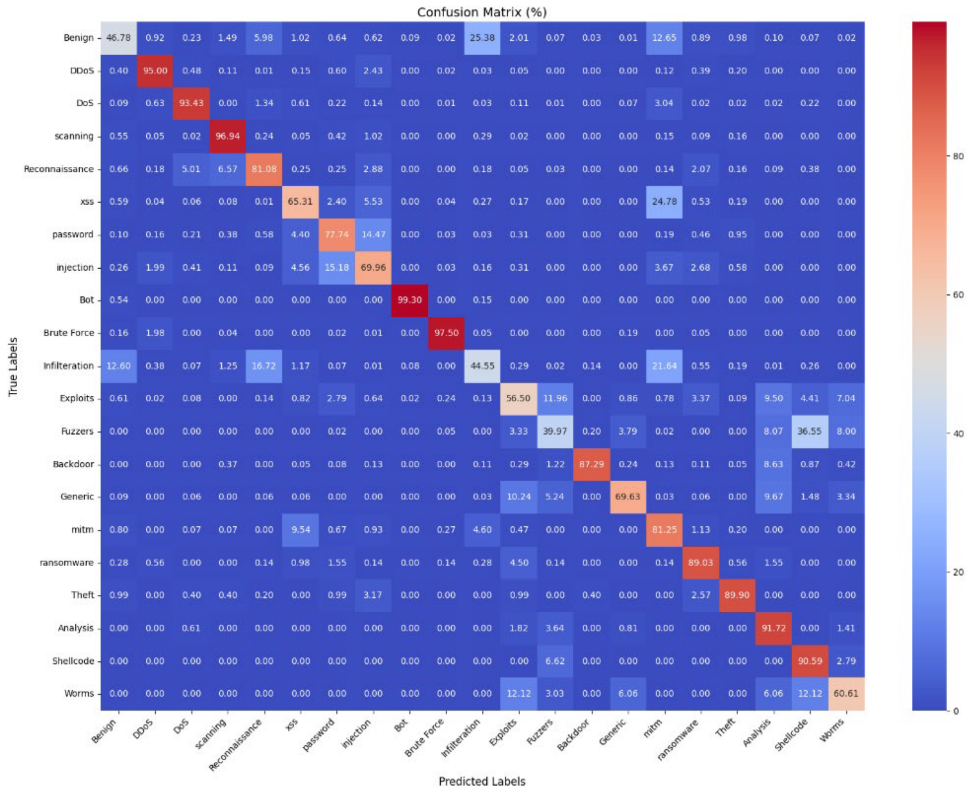
3.2.1. Mô hình CNN-LSTM



(a) Accuracy (b) Loss

Hình 10. Chỉ số huấn luyện mô hình CNN-LSTM

Độ chính xác trên tập huấn luyện của mô hình tăng dần theo số epoch và đạt giá trị cao nhất khoảng 0.75–0.80. Trên tập kiểm tra, mặc dù xuất hiện một số dao động lớn ở giai đoạn đầu, nhưng về tổng thể, cả hai đường cong đều có xu hướng tăng ổn định và gần như đạt trạng thái bão hòa sau khoảng 300 epoch. Bên cạnh đó, sự tương đồng giữa các đường cong huấn luyện và kiểm tra cho thấy mô hình không gặp tình trạng overfitting nghiêm trọng.



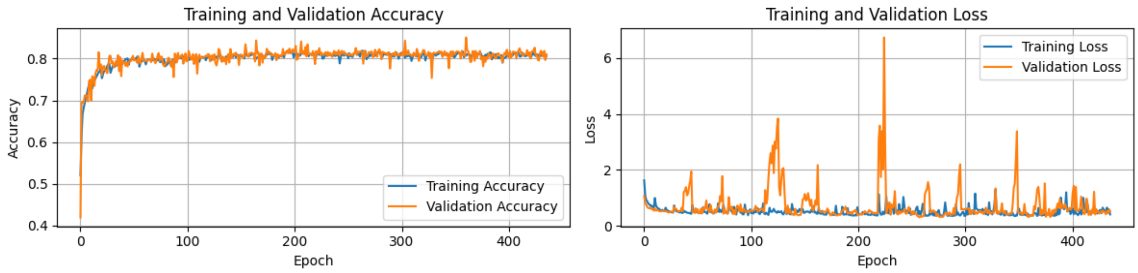
Hình 11. Ma trận nhầm lẫn mô hình CNN-LSTM

Màu sắc đậm trên đường chéo chính của ma trận nhầm lẫn cho thấy việc áp dụng kỹ thuật đánh trọng số lớp đã giúp mô hình cải thiện đáng kể khả năng nhận diện và phân loại, đặc biệt đối với các lớp thiểu số như infiltration, MITM, ransomware, theft, analysis, shellcode và worms. Tuy nhiên, riêng lớp infiltration vẫn đạt độ chính xác

ISSN: 3126-2953
E-ISSN: 3126-3097

tương đối thấp. Ngược lại, các lớp đa số chịu ảnh hưởng bởi cơ chế đánh trọng số, dẫn đến sự suy giảm độ chính xác, điển hình là lớp benign giảm từ 98,79% xuống còn 46,78% so với kịch bản trước.

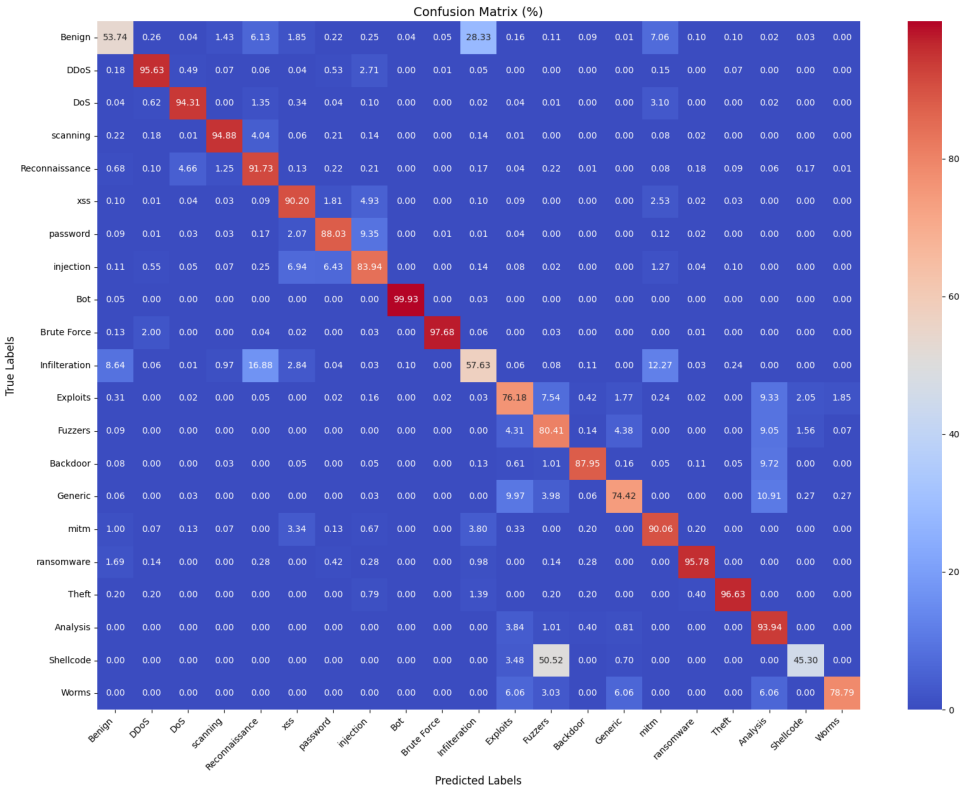
3.2.2. Mô hình RNN



(a) Accuracy (b) Loss

Hình 12. Chỉ số huấn luyện của mô hình RNN

Tương tự như mô hình CNN-LSTM, mô hình RNN cho thấy khả năng học tốt với độ chính xác trên tập huấn luyện và kiểm tra đều tăng dần và ổn định ở mức cao (khoảng 0.8). Bên cạnh đó, sự dao động và hội tụ tại các epoch cuối cho thấy mô hình không bị overfitting trong quá trình học. Đường cong độ mất mát giảm mạnh ban đầu và sau đó ổn định ở mức 0.6. Mặc dù có những đỉnh đột biến lớn trên tập kiểm tra, nhưng sự ổn định và hội tụ càng về sau cho thấy mô hình đang đạt được kết quả khá tốt.



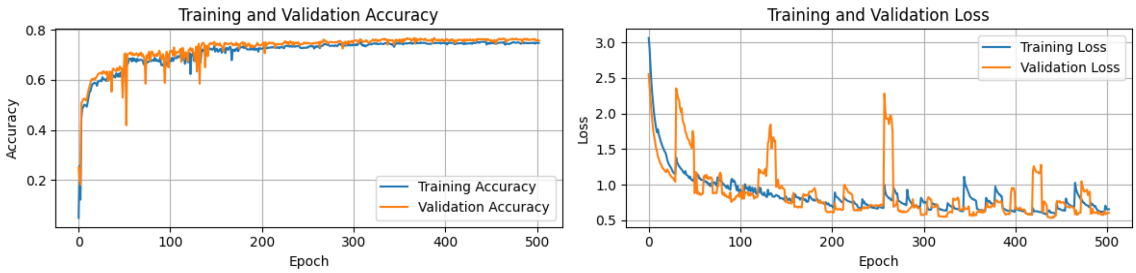
Hình 13. Ma trận nhầm lẫn của mô hình RNN

Kết quả cho thấy việc áp dụng trọng số lớp giúp cải thiện đáng kể khả năng học của các lớp thiểu số, tuy nhiên đồng thời làm suy giảm hiệu năng phân loại đối với

ISSN: 3126-2953
E-ISSN: 3126-3097

một số lớp đa số. Điều này cho thấy sự đánh đổi giữa khả năng nhận diện lớp thiểu số và độ chính xác của các lớp chiếm ưu thế trong tập dữ liệu.

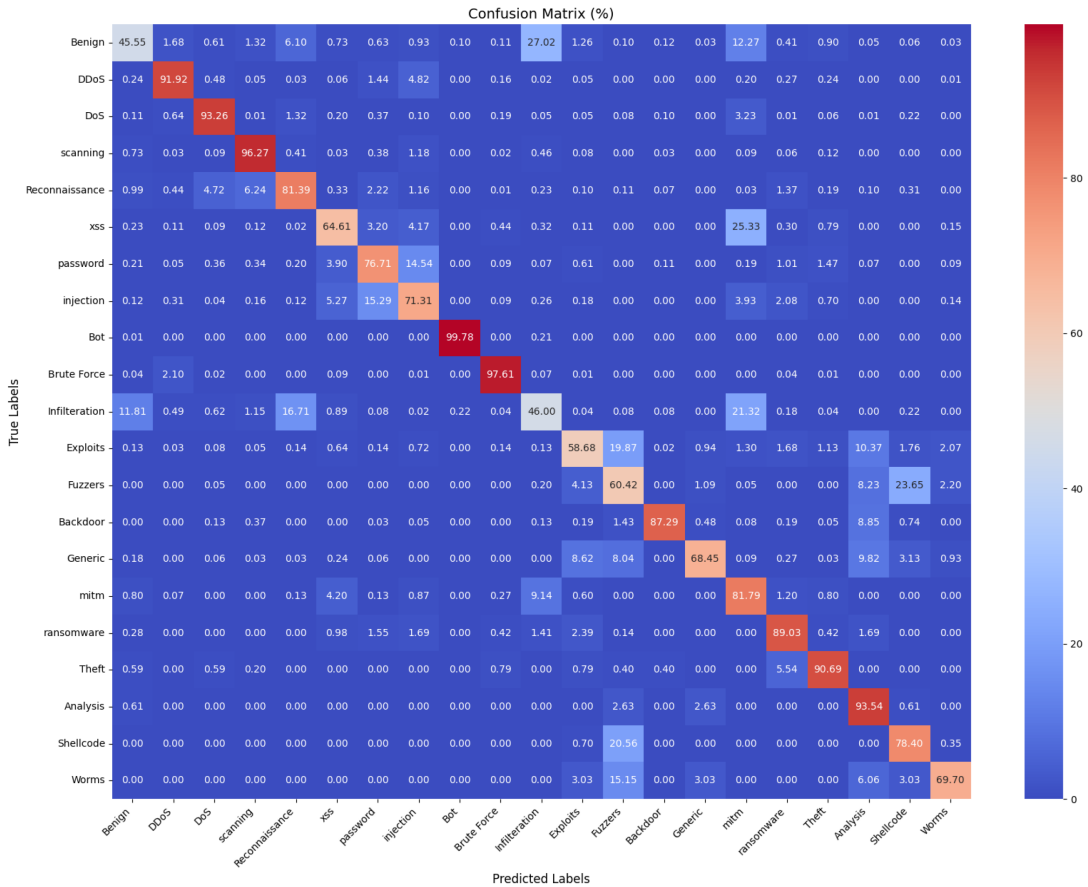
3.2.3. Mô hình LSTM



(a) Accuracy (b) Loss

Hình 14. Chỉ số huấn luyện của mô hình LSTM

Tương tự như hai mô hình trước, LSTM cho thấy khả năng học tốt với độ chính xác trên tập huấn luyện và kiểm tra đều tăng dần và ổn định ở mức khoảng 0.75. Độ chính xác trên tập kiểm tra có những dao động ban đầu lớn, nhưng sau đó dần ổn định và gần với độ chính xác trên tập huấn luyện. Về độ mất mát thì trên tập kiểm tra có nhiều biến động lớn ở khoảng đầu nhưng ổn định dần về mức thấp và gần với độ mất mát của tập huấn luyện.



Hình 15. Ma trận nhầm lẫn của mô hình LSTM

ISSN: 3126-2953
E-ISSN: 3126-3097

Tương tự như 2 mô hình trước trong kịch bản 2 này, thì trọng số đã giúp cho mô hình có thể học được tốt các lớp thiểu số mặc dù độ chính xác chỉ ở mức 70-90% nhưng trọng số cũng ảnh hưởng xấu đến quá trình học của các lớp thiểu số như benign làm giảm nhiều độ chính xác của các lớp này. Bên cạnh đó, sự xuất hiện của các điểm sáng nằm ngoài đường chéo chính cũng thể hiện là mô hình cũng còn nhầm lẫn nhiều và chưa được tối ưu cần được cải thiện thêm.

3.2.4. Tổng quan kết quả kịch bản 2

(a) Accuracy (b) Loss

Hình 16. So sánh Accuracy và Loss của các mô hình trong kịch bản 2

Bảng 4. So sánh hiệu năng các mô hình theo Macro metrics trong kịch bản 2

Mô hình	Accuracy	Macro-Precision	Macro-Recall	Macro-F1	Weighted-F1
CNN-LSTM	0,76	44,17	77,34	45,91	82,08
RNN	0,808	49,77	84,15	54,01	85,46
LSTM	0,75	40,48	78,21	42,97	80,9

Trong kịch bản 2, các mô hình được huấn luyện trên tập dữ liệu sau khi áp dụng kỹ thuật đánh trọng số lớp nhằm giảm thiểu ảnh hưởng của sự mất cân bằng dữ liệu. Kết quả huấn luyện cho thấy độ chính xác trên tập huấn luyện và tập kiểm tra của cả ba mô hình đều đạt mức trung bình (khoảng 0,75–0,81) và có xu hướng ổn định, thể hiện qua sự tương đồng giữa các đường cong Accuracy và Loss trong Hình 16. Mặc dù độ chính xác tổng thể giảm so với kịch bản 1, sự ổn định của quá trình huấn luyện cho thấy các mô hình không gặp hiện tượng overfitting nghiêm trọng khi áp dụng cơ chế đánh trọng số.

Xét theo các chỉ số đánh giá trung bình macro (Bảng 4), việc áp dụng trọng số lớp giúp cải thiện đáng kể khả năng nhận diện các lớp thiểu số, thể hiện qua giá trị Macro-Recall cao hơn so với kịch bản 1 ở cả ba mô hình. Trong số đó, mô hình RNN đạt hiệu năng tốt nhất với Macro-F1 đạt 54,01%, vượt trội so với CNN–LSTM (45,91%) và LSTM (42,97%). Tuy nhiên, sự cải thiện đối với các lớp thiểu số đi kèm với sự suy giảm độ chính xác tổng thể và Weighted-F1, phản ánh rõ sự đánh đổi giữa khả năng nhận diện lớp hiếm và hiệu năng phân loại của các lớp đa số.

Nhìn chung, kịch bản 2 cho thấy việc

ISSN: 3126-2953
E-ISSN: 3126-3097

áp dụng kỹ thuật đánh trọng số lớp đã giúp các mô hình học được nhiều lớp hơn và giảm hiện tượng bỏ sót các lớp hiếm so với kịch bản 1. Tuy vậy, mức độ nhầm lẫn giữa các lớp vẫn còn đáng kể, đặc biệt đối với các lớp cực hiếm, cho thấy cần phải tiếp tục nghiên cứu cải thiện thông qua các kỹ thuật cân bằng dữ liệu nâng cao và các mô hình học sâu hiệu quả hơn trong các nghiên cứu tiếp theo.

4. KẾT LUẬN

Nghiên cứu này đã tiến hành so sánh các kỹ thuật học máy CNN-LSTM, RNN và LSTM trong phân loại tấn công mạng, đồng thời áp dụng phương pháp đánh trọng số lớp (class-weight) để xử lý vấn đề mất cân bằng dữ liệu trên tập NF-UQ-NIDS-v2. Kết quả cho thấy, các mô hình huấn luyện trên tập dữ liệu gốc bị ảnh hưởng đáng kể bởi sự mất cân bằng, dẫn đến khả năng nhận diện kém đối với các lớp thiểu số và gây khó khăn khi triển khai thực tế. Việc áp dụng kỹ thuật đánh trọng số lớp đã góp phần cải thiện khả năng nhận diện và phân loại các lớp hiếm, giúp các mô hình học được đầy đủ cả 21 lớp tấn công trong tập dữ liệu. Đáng chú ý, trong cùng điều kiện áp dụng trọng số lớp, mô hình RNN thể hiện hiệu năng vượt trội hơn so với các mô hình CNN–LSTM và LSTM, dù các kiến trúc sau được phát triển muộn hơn.

Tuy vậy, phương pháp này vẫn còn một số hạn chế như độ chính xác chưa cao, vẫn tồn tại nhầm lẫn giữa các lớp, và độ chính xác của một số lớp đa số bị ảnh hưởng bởi việc điều chỉnh trọng số. Trong các nghiên cứu tiếp theo, có thể mở rộng bằng cách kết hợp phương pháp đánh trọng số với các kỹ thuật cân bằng dữ liệu tiên tiến hơn như SMOTE biến thể (ADASYN, Borderline-SMOTE), hoặc áp dụng các mô hình học sâu hiện đại tích hợp cơ chế Attention và các mô hình kết hợp (ensemble learning), nhằm cải thiện độ chính xác, giảm nhầm lẫn và tăng khả năng khái quát hóa của mô hình khi triển khai thực tế.

TÀI LIỆU THAM KHẢO

[1] H. Liu, B. Lang, M. Liu, and H. Yan, “CNN and RNN based payload classification methods for attack detection,” *Knowl.-Based Syst.*, vol. 163, pp. 332–341, Jan. 2019, doi: 10.1016/j.knosys.2018.08.036.

[2] H. Kamal and M. Mashaly, “Enhanced Hybrid Deep Learning Models-Based Anomaly Detection Method for Two-Stage Binary and Multi-Class Classification of Attacks in Intrusion Detection Systems,” *Algorithms*, vol. 18, no. 2, p. 69, Jan. 2025, doi: 10.3390/a18020069.

[3] R. B. Basnet, R. Shash, C. Johnson, L. Walgren, and T. Doleck, “Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks,” *J. Internet Serv. Inf. Secur.*, vol. 9, no. 4, pp. 1–17, Nov. 2019, doi: 10.22667/JISIS.2019.11.30.001.

[4] M. Radhi Hadi and A. Saher Mohammed, “An Efficient Deep Learning Approach for Network Intrusion Detection System on Software Defined Network,” *Int. J. Netw. Secur. Its Appl.*, vol. 14, no. 4, pp. 1–14, July 2022, doi: 10.5121/ijnsa.2022.14401.

[5] A. A. Hagar and B. W. Gawali, “Apache Spark and Deep Learning Models for High-

ISSN: 3126-2953
E-ISSN: 3126-3097

Performance Network Intrusion Detection Using CSE-CIC-IDS2018," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Aug. 2022, doi: 10.1155/2022/3131153.

[6] Z. Long, H. Yan, G. Shen, X. Zhang, H. He, and L. Cheng, "A Transformer-based network intrusion detection approach for cloud security," *J. Cloud Comput.*, vol. 13, no. 1, p. 5, Jan. 2024, doi: 10.1186/s13677-023-00574-9.

[7] Y.-C. Wang, Y.-C. Houng, H.-X. Chen, and S.-M. Tseng, "Network Anomaly Intrusion Detection Based on Deep Learning Approach," *Sensors*, vol. 23, no. 4, p. 2171, Feb. 2023, doi: 10.3390/s23042171.

[8] L. Ashiku and C. Dagli, "Network Intrusion Detection System using Deep Learning," *Procedia Comput. Sci.*, vol. 185, pp. 239–247, 2021, doi: 10.1016/j.procs.2021.05.025.